

Conceptualizing and measuring perceived service complexity

Article (Accepted Version)

Kostopoulos, Ioannis, Boukis, Achilleas and Lodorfos, George (2019) Conceptualizing and measuring perceived service complexity. *Journal of Marketing Theory and Practice*, 27. pp. 38-54. ISSN 1069-6679

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/76313/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.



Conceptualizing and Measuring Perceived Service Complexity

Journal:	<i>Journal of Marketing Theory and Practice</i>
Manuscript ID	MMTP-16-12-1081.R2
Manuscript Type:	Original Article
Keywords:	Perceived service complexity, service employees, scale development, frontline employees
Abstract:	<p>This study focuses on the notion of Perceived Service Complexity (PSC). PSC captures 'the difficulty to assimilate the service delivery process, as perceived by frontline employees (FLEs)' and is conceptualized through the development and validation of a multidimensional construct consisting of three factors (Task-Related, Customer-Derived, 'Service Nature'-Derived Complexity). The findings add to the organizational frontline literature in how aspects of FLEs' working environment shape their ability to assimilate service delivery and perform their roles during the service encounter. Managerial practice can be informed of the distinct elements that shape PSC and of its ramifications for designing service delivery systems for different types of service.</p>

SCHOLARONE™
Manuscripts

Conceptualizing and Measuring Perceived Service Complexity

Abstract

This study focuses on the notion of Perceived Service Complexity (PSC). PSC captures ‘the difficulty to assimilate the service delivery process, as perceived by frontline employees (FLEs)’ and is conceptualized through the development and validation of a multidimensional construct consisting of three factors (Task-Related, Customer-Derived, ‘Service Nature’-Derived Complexity). The findings add to the organizational frontline literature in how aspects of FLEs’ working environment shape their ability to assimilate service delivery and perform their roles during the service encounter. Managerial practice can be informed of the distinct elements that shape PSC and of its ramifications for designing service delivery systems for different types of service.

Keywords: Perceived service complexity, frontline employees, scale development

1. INTRODUCTION

Today, service industries are becoming increasingly competitive and unpredictable due to the increased variety of services offered, the ongoing introduction of radical innovations and, most importantly, more sophisticated and diversified customer needs and preferences (Brooker *et al.* 2012; Chen *et al.* 2014). In such a dynamic environment, service providers struggle to optimize their service delivery process, which remains a key determinant of their market performance (Iyer *et al.* 2014). The design and implementation of service procedures constitute the key components of a successful service delivery process (Teixeira *et al.* 2012), especially for firms whose frontline employees (FLEs) are directly involved in the delivery, promotion and sale of service offerings to consumers (Clark *et al.* 2000; Kostopoulos *et al.* 2012). A critical but relatively unchallenged aspect of successful service design and implementation is associated with the understanding of complexity's impact on the service delivery process, which is rooted in service interactions (Shostack 1987) and has important implications for the experience of both sides of the service encounter (i.e. FLEs and customers) (Braun and Hadwich 2016; Mikolon *et al.* 2015).

Firm-wise, service complexity enables operational efficiency in the service design process (Jan Angelis and Thompson 2007), through adjusting service delivery according to the degree of complexity of the service delivered (e.g. Wang *et al.* 2014). Regarding customers, high service complexity can hamper their service experience, due to the increased cognitive effort required on their behalf to complete such interactions (Holm *et al.* 2012; Mikolon *et al.* 2015). Nevertheless, the impact of service complexity on the FLEs' side still remains unchallenged, despite that they strive to balance between conforming to standardized role requirements and managing the variation of customer needs (Aksin and Masini 2008). In such circumstances, the outcome of the

service encounter can be affected, as FLEs’ ability to perform can be compromised from complex service offerings. This is due to the increased effort required which results in greater cognitive demands from individuals, reducing their capacity to info-processing (Vohs *et al.* 2008). Nevertheless, service complexity is viewed either from an intra-organizational perspective where its benefits and costs in internal exchanges are assessed (e.g. Braun and Hadwich 2016) or is treated as an operational feature of service delivery assessed through objective proxies, such as the number of intermediate steps (e.g. Martínez-Tur *et al.* 2001).

Hitherto, some challenges for organizational frontline emerge from this discussion. First, the lack of knowledge about FLEs’ perceived complexity prevents line managers from ascribing lower service performance to work overload or other contextual factors, which is vital in performance-based services. Second, high levels of service complexity might require the development of more detailed job descriptions and more customized training on the tasks that FLEs need to undertake. As a result, service organizations cannot accurately act upon the negative outcomes of FLEs’ perceptions of service complexity without scrutinizing their perceived job demands due to service complexity and thus its impact on their service delivery efforts needs to be charted (Braun and Hadwich 2016; 2017).

Echoing these challenges, this study aims to expand prior conceptualizations of service complexity and provide an exhaustive view of perceived service complexity for the service encounter reality, by introducing an FLE-based conceptualization of complexity, namely Perceived Service Complexity (PSC). PSC captures ‘*the difficulty to assimilate the service delivery process, as perceived by FLEs*’. Drawing on the job demands-resources and the job characteristics frameworks, this study advances the service management literature in shedding light on how aspects of FLEs’ proximal working environment shapes their ability to assimilate

the service delivery process and perform successfully their roles during their interactions with customers. On this basis, three key objectives are set: a) to develop a comprehensive conceptualization of PSC and identify its main underlying dimensions, b) to develop and empirically test a parsimonious, valid and reliable scale to measure PSC and c) to assess the criterion validity of a formative model to describe PSC and its impact on its two well-established consequences, (i.e. role clarity and job performance).

The next section presents the conceptualization of PSC and the extensive literature review on which it is based. The sections that follow present the formative model developed to measure PSC, as well as the research design and data analysis for the two studies conducted to test the model's validity. The final section offers a discussion of the findings and some directions for future research.

2. LITERATURE REVIEW

2.1 From Service Complexity to Perceived Service Complexity

Service complexity is a characteristic related to the success of a new service (De Brentani 1989) as well as the sustainability of an existing one (Surprenant and Solomon 1987; Danaher and Mattsson 1998). Existing definitions of service complexity mostly adopt an operational view of the construct's meaning (e.g. Silvestro *et al.* 1992; Kreye *et al.* 2015). Work in the operations management literature assesses the complexity of a service on the basis of process outcomes differentiating between service complicatedness and difficulty (Soteriou and Chase 1998; De Castro Lobo *et al.* 2010; Kreye *et al.* 2015). Work in organizational behaviour mostly focus on the complexity of the tasks employees have to carry out and rarely explore service complexity from the service provider's point of view (Chen *et al.* 2001; Braun and Hadwich 2016). In a similar vein, the marketing literature often treats service complexity as an objective service

attribute, which remains the same for a given service, regardless of employees’ perceptions of it (e.g. Shostack 1987; Braun and Hadwich 2016). Recent work also considers service complexity from a customer perspective and explores its impact on customers’ experience with the firm (Mikolon *et al.* 2015; Balaji *et al.* 2017).

The aforementioned conceptualizations rarely address the impact of service complexity on FLEs’ ability to perform during the service encounter, which largely determines customers’ view of the firm, especially in performance-based services (Gounaris and Boukis 2013). FLEs’ ability to perform is not only affected by tangible role determinants, such as structural job characteristics, but it is also shaped from intangible ones, such as the mental effort required on their behalf when delivering the service (Gillison *et al.* 2016). Capturing these determinants of service complexity is vital, as it remains individually experienced; the same level of objective complexity within a service process may have a varied mental and psychological effect on different FLEs. For instance, the actual number of intermediate steps in a service delivery process, which is a determinant of its complexity (Shostack 1987), does not precisely capture the degree of complexity that each FLE perceives. Hence, its actual impact on the delivery process cannot be accurately assessed without considering both tangible and intangible aspects of the service delivery process that determine FLEs’ perceptions of complexity.

In understanding service complexity, prior research has utilized the cognitive capacity framework (Lalwani 2009) to explain how customer value is affected from service complexity (Mikolon *et al.* 2015). Yet, a solid theoretical understanding of how FLEs experience service complexity is still missing from the literature. This is of paramount importance for the accurate measurement of perceived complexity and the identification of managerial practices to deal with it. The present study advances an FLE-based conceptualization of service complexity which caters for FLEs’ perceived difficulty to assimilate the service delivery process. Drawing on the

1
2
3 Job Demands-Resources framework (Demerouti *et al.* 2001), PSC is viewed as a job demand
4 which impairs FLEs' performance, due to the higher levels of cognitive effort required on their
5 behalf to meet role requirements (Bakker *et al.* 2005). Next, the theoretical underpinning for this
6 construct is analytically discussed and each of its three underlying dimensions is established
7 (Figure 1).
8
9
10
11
12
13
14

15 Place Figure 1 about here
16
17

18 **2.2 Identifying the Dimensions of PSC**

19

20
21 Despite several studies in the literature conceptualizing service complexity from an operational
22 viewpoint (e.g. Aksin and Masini 2008) or an internal supplier perspective (e.g. Braun and
23 Hadwich 2016; 2017), none of them incorporates FLEs' views, despite that they are the key
24 stakeholders affected from service inseparability during service delivery. Following an extensive
25 review of the concept across the OB, services and operation literatures, three major dimensions
26 underlying PSC emerge: *Task-Related Complexity*, *Customer-Derived Complexity* and *Service*
27 *Nature-Derived Complexity*. These three dimensions represent the input of the three main
28 sources of complexity, as conceptualized in the extant literature, for FLEs; first, the way the
29 service is designed by the service provider (Chase and Tansik 1983); second, the input from
30 external participants (i.e. the customers) (Dagger *et al.* 2009) and, third, the effect from the
31 nature of the service itself, which cannot be adjusted by the service provider or customers
32 (Laroche *et al.* 2001; 2004).
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 **2.2.1 Task-Related Complexity**

50

51
52 Task-related complexity captures FLEs' perceived complexity of the task in hand, which remains
53 a key source of complexity in service delivery. The aspects of task-related complexity should not
54
55
56
57
58
59
60

only be measured objectively (e.g. actual duration of the service delivery), but also by capturing FLEs’ views of the process (e.g. how long they perceive the duration of service delivery). The theoretical underpinning of this factor lies on job design theory (Hackman and Oldman 1976), which suggests that employees can be motivated through the optimal design of their jobs along five elements (i.e. variety, identity, significance, autonomy, feedback). As organizations strive to encourage high job motivation through enhancing jobs along these elements, task-related complexity is also affected. For example, increased job variety increases the complexity of the service as a range of sub-processes must be designed and undertaken, which in turn require a variety of additional resources (Chase and Tansik 1983).

Traditional marketing or operations management literature focus on the *number of tasks* involved in the delivery of the service and the *difficulty* of their execution (e.g. Andaleeb and Basu 1995; Germain *et al.* 2001; Martínez-Tur *et al.* 2001). This work is in line with definitions of *task complexity* in the pertinent literature that consider a job task complex when it involves a large number of difficult steps (e.g. Chen *et al.* 2001). In general, service delivery processes that comprise several difficult tasks (e.g. hotel accommodation) are perceived as more complex by all parties involved than those which require fewer and easier tasks (e.g. ticket purchase) (Kostopoulos *et al.* 2012). In fact, this is the theoretical argument with regard to a system’s complexity: complex systems consist of many elements that interact with each other in ways that heavily influence the probability of later non-predictable events (Amaral and Uzzi 2007).

What is also important is the extent to which discrete steps during service provision are different from each other (i.e. task variety) (Lightfoot and Gebauer 2011); in a service setting, intermediate steps often differ significantly, and additional sub-processes must be designed and implemented, which in turn require a greater variety of organizational resources and employee

skills (Chase and Tansik 1983; Kreye *et al.* 2015). Thus, diverse employee training and development processes need to be adopted, and different equipment acquired, which increases FLEs' perceptions about the complexity of the service (Devlin 2001). Taking flight services as an example, the tasks involved are so disparate that a significant number of varied tasks need to be completed by staff (e.g. customer service, safety control, check-in), and this adds further complexity to the service.

Another task-related aspect that determines service complexity is the duration of the execution of the tasks. Pertinent literature is replete with studies that consider the duration of each intermediate step in the service delivery and the overall waiting period for the customer to be major indicators of complexity (e.g. Rafiq and Ahmed 1998; Holm *et al.* 2012). Although *duration* is related to each task's difficulty, diversity and interdependence with other tasks is not solely determined by these three factors and hence it stands on its own as a unique task attribute (Silvestro *et al.* 1992); the more time FLEs spend on a task, the lower their cognitive alertness becomes, which increases their perception of complexity (Mikolon *et al.* 2015). The above discussion suggests that task-related complexity can be determined by FLEs' perceptions of four elements related to the tasks FLEs need to carry out: *a) the number of tasks, b) the difficulty of tasks, c) the variety of tasks and d) the duration of tasks.*

2.2.2 Customer-Derived Complexity

Customer-derived complexity is the complexity arising from customers' participation in the service delivery (Rafiq and Ahmed 1998; Mikolon *et al.* 2015). In principle, the delivery of a service becomes more challenging for FLEs when *customers participate* heavily in the process (Dagger *et al.* 2009; Dong *et al.* 2015); hence, services involving more intense customer

participation, such as health care or education, are more complex than those in which low interaction exists between FLEs and customers, such as fast-food outlets or car repair services. Therefore, the intensity of service interactions should be considered a determinant of customer-derived complexity.

Customers’ participation also adds to the complexity of a service because their behavior is often *unpredictable* (Surprenant and Solomon 1987), which increases the heterogeneity of the service. Low predictability of customer behavior makes it difficult for service providers to plan and execute the service delivery process (Hjort *et al.* 2013). When FLEs have a less clear picture on what to expect during the service encounter, it enhances their uncertainty, making the service appear as more complex. Hence, service interactions become more perplexing, which requires FLEs to display a mix of task-, relationship-, and self- focused behaviors (Bradley *et al.* 2013) which increase the degree of service complexity FLEs perceive.

Another source of complexity is the degree to which the service offer can be customised (Silvestro *et al.* 1992). Holm *et al.* (2012), regard service complexity as a function of ‘*the degree of variation in service needs and requirements that invoke differential activities on an organization across customer-facing functions*’ (p. 394). As more options (alternative scenarios) become available to customers, a wider variety of actions are included in the service delivery. As a result, the service plan includes more parameters, which increase the complexity of the service significantly (Rafiq and Ahmed 1998). This is why services such as car registration, which involve specific and predictable customer actions, are considered simple, whereas services such as legal advice services are considered more complex (Buckley 2003). Therefore, the variation of customer needs is expected to influence FLEs’ perceived service complexity.

Another source of customer-derived complexity pertains to the simultaneous presence of many customers (Holm *et al.* 2012). When FLEs deal with many customers at once, they find it more difficult to execute the tasks involved in service delivery (Hoffman and Turley 2002; Ng *et al.* 2007). For example, therapists view group therapy as a more complex process than one-to-one consultation. Similarly, a bartender will find it much easier to deal with one customer's order than multiple customers' requests at the same time. The simultaneous presence of many customers impairs FLEs' ability to predict customers' potential behavior and depletes their resources quicker (Singal 2008). In light of the above discussion, customer-derived complexity can be determined from: *a) the intensity of customer participation, b) the predictability of customer behavior, c) variation in customer needs and d) the simultaneous presence of multiple customers during the service encounter.*

2.2.3 'Service Nature'-Derived Complexity

The marketing paradigm suggests that services are axiomatically more difficult to grasp than products due to their intangible nature (Shostack 1987; Lovelock 1983). This study asserts that PSC is also determined by FLEs' ability to understand the service, which derives from the imprecise and intangible nature of service interactions (Simon and Usunier 2007). Thus, the third dimension of PSC is 'Service Nature-Derived' Complexity, which reflects the cognitive difficulty that the nature (type) of the service poses for FLEs.

The first source of 'Service Nature-Derived' Complexity is mental intangibility, or the extent to which a service is difficult to grasp mentally (Laroche *et al.* 2001; 2004). Mental intangibility can impair cognitive understanding and generate associated difficulties for all individuals involved in a service encounter. Hence, it should be considered a determinant of complexity

(Devlin 2007). For instance, higher education is a more mentally intangible service than a business loan; therefore, it is viewed as more complex by the staff involved, although the individual tasks required are not necessarily more difficult.

In addition to mental intangibility, there is the generality of a service, which derives from its nature and is determined by its *obscurity*; that is how general and/or specific an individual perceives a particular service (Laroche *et al.* 2001). More abstract services are those that cannot be easily identified by precise definitions, features and/or outcomes (Laroche *et al.* 2004). Therefore, it is reasonable to assume that abstract services are viewed as more complex by FLEs: for example, psychotherapy, which is *the process of dealing with a person's mental or emotional problems through conversation*, is a more abstract service than a haircut.

Another source that underlies the nature of a service pertains to the amount of knowledge required by FLEs to fully understand the nature of that service (Andaleeb and Basu 1995; Devlin 2007). Services for which FLEs need significant knowledge or intellectual capital in order to fully understand them (e.g. technologically advanced or medical services) will be perceived as more complex than services which do not carry such a requirement. Therefore, the PSC of the former will be higher.

Finally, FLEs' inability to get a complete overview of the service process as a whole is another source of 'service nature'-derived complexity (Swanson and Kelley 2001). In many cases, FLEs have an explicit idea only about the phase of service delivery they are involved in, being unaware of other important components of the service offering (Lings and Brooks 1998). Often, FLEs have low visibility or limited understanding of some aspects of the services process which makes it more challenging for them to fully understand the service. For example, in the case of air transportation, ground staff do not have a clear picture of the in-flight service and vice

versa for flight attendants. In summary, PSC can also be determined by the nature (type) of the service, which is reflected on four elements: *a) mental intangibility, b) generality, c) knowledge requirements and d) incomplete overview.*

2.3 A Formative Model of Perceived Service Complexity

This study conceptualizes Perceived Service Complexity (PSC) as a multidimensional, second-order construct with three formative, first-order factors (*Task-Related Complexity, Customer-Derived Complexity* and *'Service Nature'-Derived Complexity*).

As such, the model is a reflective one at the first order, as three latent factors emerge, which are reflected upon their indicators. The first latent factor is task-related complexity, which is reflected upon the number of tasks, the difficulty of tasks, the variety of tasks and the duration of tasks; the second one is customer-derived complexity, which is reflected upon the intensity of customer participation, the predictability of customer behavior, the variation in customer needs and the simultaneous presence of multiple customers and the third factor is 'service nature'-derived complexity, which is reflected upon mental intangibility, generality, knowledge requirements and incomplete overview. The reason for this level of the model being reflective is that each complexity dimension is an underlying concept that has an effect on its indicators (Bollen and Lennox 1991); in other words, each latent (complexity) factor determines its indicators and not vice versa (Diamantopoulos *et al.* 2008). In contrast, at the second order, the three factors form, rather than reflect, the overall construct of PSC. As this is the first time that PSC is conceptualized as a multidimensional, second-order construct, some arguments are presented to support the view that the model is a formative one at the second order.

2.3.1 Covariance between the Indicators

In this theoretical conceptualization, the indicators for the three factors of PSC are not necessarily related to each other: for instance, some services are considered complex because of the difficulty of the associated tasks, while others are complex as they involve high customer participation and multiple service scenarios from which customers can choose (Silvestro *et al.* 1992). Thus, some services with high task-related and low customer-derived complexity (e.g. technical support, public prosecution) are equally complex as services with low task-related and high customer-derived complexity (Buckley 2003). Similarly, many services are complex due to their nature, even though their task-related and customer-derived complexity is low (e.g. financial services) (Devlin 2007). It becomes evident, therefore, that the significance of the correlations between the three factors of complexity cannot, theoretically, be predicted: a reflective model would assume correlations between indicators, whereas such an assumption cannot be made for a formative model (Law *et al.* 1998; Diamantopoulos *et al.* 2008). This signifies that overall, PSC can be better explained by a formative than a reflective model.

2.3.2 Direction of Causality from Construct to Factors

The direction of causality in a second-order, formative model moves from factors to construct, whereas in reflective models the opposite is evident (Diamantopoulos and Winklhofer 2001). Following the conceptualization of PSC in this study, changes in FLEs' perceptions of the three factors (task-related complexity, customer-derived complexity and 'service nature'-derived complexity) lead to changes in their overall perceived complexity: when one of the complexity factors increases, so does the overall complexity of the service. However, when FLEs' perceptions on the overall service complexity change, this does not necessarily mean that their

perceptions of any other factor will change; in fact, it is just as likely that there will be a change for only one or two of the factors, since the three are not necessarily interrelated. Consequently, the direction of causality moves from the three factors to the overall variable (Diamantopoulos *et al.* 2008).

2.3.3 Interchangeability of the Indicators

In a reflective model, the indicators have similar content and therefore, construct validity will remain the same if a single indicator is eliminated, although the reliability of the construct will suffer (Jarvis *et al.* 2003). In contrast, in a formative model, each factor is only a component of the whole, and the whole becomes incomplete if any components are omitted (Lin *et al.* 2005).

In this PSC model, the three factors are theoretically distinct; hence, disregarding one of the three components will change the content validity of the overall service complexity. For instance, if customer-derived complexity is eliminated, the interpretation of the overall complexity and the predictability of the model, in general, will change: the new 'complexity' variable will overlook the 'customer participation' factor, and hence services such as nursing or child protection will be viewed as simple, contrary to how they are commonly perceived (Buckley 2003). Similarly, if task-related and/or 'service nature'-derived complexity are disregarded, then services that are actually quite complex (e.g. financial services) may also be viewed as simple, due to the standardization of the outcome.

2.4 The PSC Scale's Criterion Validity

In order to test the PSC scale's criterion validity, this study followed the recommended process (Churchill 1979; Coltman *et al.* 2008) and examined whether the construct predicts some criterion measures as it is expected to. For that reason, three research hypotheses were developed

within the formative model to test the direct impact of PSC on two variables which, in theory, are direct consequences of PSC (Chung and Schneider 2002; Kauppila 2014). These two variables are role clarity and job performance. They were selected as they constitute the most immediate consequences of high complexity for FLEs while they can remain pivotal in defining customers' experience with the service encounter (Whitaker, Dahling and Levy 2007).

The complexity of a service process has been negatively associated with the degree to which FLEs have a clear picture of their role in it (Chung and Schneider 2002). This is due to the fact that increased complexity leads to higher role conflict and ambiguity and creates confusion among staff, especially in customer-contact posts (e.g. Kauppila 2014). This in turn decreases their ability to serve individual customer needs and may hinder the success of the service provision (Hartline and Ferrell, 1996). Moreover, increased PSC may lead to increased active and latent errors made by FLEs during the service delivery, reducing their performance. On the contrary, when PSC is low, FLEs feel that they have a clear picture on what they are supposed to do and how to do it and therefore they are more likely to perform better (Whitaker *et al.* 2007). The above discussion implies that there is a negative influence of PSC on job performance, which is both direct and indirect through the decrease on FLEs' role clarity. Hence, we formulate the following research hypotheses:

H1: *PSC has a negative effect on FLEs' role clarity.*

H2: *PSC has a negative effect on FLEs' job performance.*

H3: *Role clarity has a positive effect on FLEs' job performance.*

3. RESEARCH METHOD AND DATA ANALYSIS

In order to develop and empirically test a measurement scale for PSC, the recommended scale development process is followed (Anderson and Gerbing 1988; MacKenzie *et al.* 2011). The construct's domain is determined based on an extensive literature review of the notion of complexity in the management, operations and marketing literature which, coupled with a number of interviews, was used to create an initial pool of items (Bigné *et al.* 2002). Two studies were then conducted – study 1 to validate and reliability test the initial factors, and study 2 to empirically test the validity of the formative model that captured the PSC scale – in addition to the hypotheses that were developed to test the scale's criterion validity.

3.1 Item Generation and Content Validity

The first stage is to review the existing literature and establish the PSC construct's domain (e.g. Shostack 1987; Silvestro *et al.* 1992; Laroche *et al.* 2004). The item development for the new construct was based on existing work around the notion of complexity in three relevant disciplines (i.e. management, operations and marketing). As explained in section 2, three major factors, each with four sub-factors, were identified (12 items in total). Following the review, 10 interviews with executives from several service firms (hotels, restaurants) and 13 interviews with management and marketing academics was conducted, leading to the creation of an initial pool of items to capture the underlying elements of each dimension. For each of the 12 elements, three possible items was developed and each item is assigned a Likert-type scale with anchors 1-7 (1=Totally Disagree – 7=Totally Agree).

The list of items, together with the study's overall subject and research objectives, were then given back to the same group of executives and academics who were asked to rank the items based on the degree to which they believe the items are measuring what they intend to (i.e. content validity). With the use of Q-sort tests, the item with the best content validity is selected

for each of the 12 elements and is included in the final research instrument (Jinbo *et al.* 2017). Some items' wording was slightly amended based on the experts' suggestions. At the end of this process, 12 items were included (see Table 1) and 24 were excluded from the final questionnaire. Some indicative items that were excluded from the questionnaire are reported below, as they had lower content validity and/or were not adequately categorized in any dimension by at least two-thirds of the participating experts (Malhotra 1981) are: *'The tasks we have to execute in order to deliver the service vary'*, *'The duration of the tasks involved in the service delivery is lengthy'*, *'Customers' preferences vary and hence there are several alternative ways of serving them'* and *'The service we provide to the customer is very general'*.

3.2 Study 1: Factors' Validity and Reliability Testing

3.2.1 Method

The first study is carried out in order to test the dimensionality of the 12 items (Diamantopoulos and Siguaw 2006). In doing so, four research assistants were employed in major cities in the UK (i.e. London and Leeds) (two in each city) and they initially approached a convenience sample of participants from the aforementioned cities. A restriction was additionally imposed, to draw from participants who work in both low- and high-complexity service providers. Overall, they contacted 319 participants who were eligible for participation in the study and 150 of them finally agreed to participate in the study (the response rate was around 47%). To ensure high control of the sample and accurate screening, research assistants gave each participant a hard copy of the questionnaire and remained present during questionnaire completion, which included the twelve items of the PSC scale. Moreover, participants were asked to complete the questionnaire having in mind one of the services they participate in and to indicate this type of

service. The types of service organizations where the participants were employed vary: approximately 25% of the participants worked in bars and restaurants, 20% in hotels, 15% as teaching personnel in universities, 15% in public services, 10% in banks and 15% for other service providers (e.g. consulting, personal training).

A range of sectors with varying levels of complexity were selected to ensure that different types of services would be considered. Moreover, based on a previously validated service categorization (Danaher and Mattsson 1998), t-test analysis is conducted between respondents from high and low complexity services, which is based on the participants' responses to a single item asking them to rate the extent to which they believe their job is complex. Results indicate that significant differences exist between low and high complexity service jobs ($t=-20.512$, $p<0.001$). With regard to demographics, 36% of respondents are men and 64% were women, and the average age is 34.4 years.

3.2.2 Scale Purification

The first stage of the analysis is to examine whether the PSC scale needed purification, for which the average corrected item-to-total correlations were calculated for all 3 factors and 12 items of the scale. The results showed that no item-to-total correlation is below 0.50 (see Table 1), and thus all items were suitable for inclusion and there is no need for scale purification.

Place Table 1 about here

3.2.3 Factors' Validity and Reliability

To test the unidimensionality, validity and reliability of the three factors in the PSC scale, confirmatory factor analysis (CFA) is applied using AMOS 22 software. Tables 2 and 3 present the results of the analysis, which demonstrate each factor's psychometric qualities. For all three factors, the items' loadings are more than 0.60 (Bagozzi and Yi 1988), the pertinent fit indices

are within the suggested limits (Byrne 2006) and the percentage of their explained variance (average variance extracted, AVE) is higher than 50% and higher than the maximum squared correlation between the three factors (Fornell and Larcker 1981). The three factors are also examined for internal consistency, as reflected by construct reliability, which is assessed through composite reliability (CR) and Cronbach’s alpha coefficients. For all three factors, both the composite reliability (Fornell and Larcker 1981) and Cronbach’s alpha coefficients (Nunnally 1978) are again substantially high (>0.7 each). These results indicate that all factors have adequate reliability and discriminant validity.

Place Table 2 about here

3.3 Study 2: Testing the Formative Model for PSC

3.3.1 Method

In order to verify the properties of the PSC scale generated in study 1 and test the second-order formative model, study 2 was carried out. Following the approach of study 1, five research assistants were used to reach participants from the same service sectors. To increase speed and capitalize on the personal networks of the research assistants, a snowballing sampling technique was used. Research assistants initially contacted all participants from study 1 and asked for referrals on other participants who were eligible for participation in study 2. Overall, through referrals, 619 FLEs were contacted and asked to complete the study’s questionnaire, out of whom 244 agreed to do it (response rate: 39.4%). In all cases the completion of the questionnaire took place via face-to-face interactions with members of the research team. The sample is consistent with the demographic characteristics of study 1: 25.4% of the participants worked in bars and restaurants, 20.9% in hotels, 11.8% in universities, 13.9% in public services, 10.6% in banks and 19.2% for other service providers (e.g. retail banking, personal training); the

demographic profile of the respondents is 33.2% men and 66.8% women, with an average age of 32.13 years. Again, a t-test analysis took place based on the participants' rating on their perceived job complexity; results suggest that significant differences exist between low and high complexity service jobs ($t=-24.182$, $p<0.001$).

All participants completed a questionnaire that included the twelve items of the PSC scale, two items that captured the overall PSC, four items that measured FLEs' role clarity, four items that captured their job performance and some info regarding their demographic profile. In order to capture the latter two constructs, an adaptation to the scales developed by Singh (2000) was employed. Specifically, the items used in the role clarity scale capture the degree to which FLEs have a clear picture about their role (i.e. *'How they are expected to handle the non-routine activities of the job'*, *'Which tasks they should give priority to'*, *'How they are expected to interact with the customers'* and *'How they should behave while on the job'*). The items used in the job performance scale describe the degree to which FLEs are able to perform (i.e. *'Consistently follow up on promises made to the customers'*, *'Overall, consistently provide prompt service to all customers'*, *'Provide accurate information to the customers'* and *'Perform their job reliably and accurately'*). Two Likert type items (1=Totally Disagree – 7=Totally Agree) were used to capture FLEs' overall PSC in the formative model and they refer to the degree to which: *"The service they offer to our customers is simple/complex"* and *"The delivery of the service offering is a complicated process"*.

3.3.2 Data Analysis

Construct validation: To confirm the hypothesised structure, the scale for the PSC construct should exhibit properties of a reflective first-order, formative second-order model comprised of three first-order factors: task-related complexity, customer-derived complexity, and 'service

nature'-complexity. Specifically, each item is forced to load on its intended factor and not allowed to cross-load on other factors. Also, the two items measuring overall PSC were loaded to an overall PSC factor, and two paths emanating from this second-order construct were added to the model (Bollen and Davis 2009). Finally, a path model is developed including three regression paths from the three first-order latent factors to the overall PSC latent factor.

Measurement model: In order to test the validity and reliability of the dimensions, the Anderson and Gerbing (1988) method for scale development is followed: first, unidimensionality was assessed, then both convergent and discriminant validity were determined, and finally, reliability of the scale items is evaluated. The three factor solution was initially tested providing a good fit, as indicated from the following indices ($\chi^2=224.65$, Df=51; CFI=0.921; GFI=0.919; TLI=0.916; RMSEA=0.057). Next, the unidimensional model for the 12 items was tested resulting in a chi-square of 354.693, 54 degrees of freedom, indicating relatively poor fit to the data (CFI=0.899; GFI=0.904; TLI=0.892; RMSEA=0.058). Last, the second-order solution with a reflective first level and formative second level showed a good fit with the data ($\chi^2=514.013$; Df=74; CFI=0.929; GFI=0.926; TLI=0.905; RMSEA=0.059). Based on the second-order solution which provided a better fit than the other solutions, a complete list of the 12 items, with the factor loadings for each item is provided in Table 3.

Place Table 3 about here

Preliminary support for convergent validity is found, given that all items loaded highly and significantly on their specified constructs. Moreover, the average variance extracted (AVE) for each construct exceeded 0.50. Following this, a formal evaluation is made of the discriminant validity of the organizational culture profile (OCP) dimensions, using the method outlined by Fornell and Larcker (1981), by comparing the AVE to the squared correlations between the items

included. All AVE values exceeded the squared correlations for each pair, thus displaying adequate discriminant validity (see Table 4).

Following these assessments for validity, the reliability of the scales for each OCP dimension is determined. Reliability is assessed by calculating the construct reliability based on the standardized factor loadings and error variances, as well as Cronbach's alpha: the estimates from both calculations exceeded 0.70 for all dimensions, ranging from 0.728 to 0.839. These results thus suggest that the PSC dimensions meet the requirements for construct reliability. Table 4 provides the AVEs, reliability estimates and correlations for the PSC indicators.

Place Table 4 about here

Formative model: With regard to the second order of the formative model, the standardized estimates of each first-order factor compared to the second-order factors were found positive and significant (see Table 5). This, together with the good overall fit of the two-level mixed model, confirms the hypothesized structure of the scale.

Place Table 5 about here

3.3.3 PSC's Criterion Validity

Before testing the three research hypotheses for PSC's criterion validity, the measurement model was established including the second-order mixed model, which described PSC, along with the influence of PSC on role clarity and job performance; results indicate a good fit of the model (i.e. $\chi^2=367.95$; Df=172; CFI=0.925; TLI=0.908; RMSEA=0.068). Having established the measurement model, the structural model of PSC was assessed, where the construct validity and reliability of the scales used to measure role clarity and job performance were estimated. Table 6 displays the constructs' validity and reliability along with the intercorrelation matrix.

Place Table 6 about here

These results reveal that the model fits the data well ($\chi^2=860.01$; Df=224; CFI=0.926; TLI=0.909; RMSEA=0.065). Moreover, every proposed path in the model is statistically significant (see Table 7): Perceived Service Complexity has a negative influence on FLEs' role clarity (-0.714), confirming H1; similarly, H2 is verified by PSC having a negative influence on FLEs' job performance (-0.543); and finally, role clarity has a positive influence on FLEs' job performance (0.426), confirming H3. Thus, the explanatory power of the model remains quite satisfactory, as 49.2% of the variance of PSC, 50.6% of role clarity and 77.6% of job performance are explained.

Place Table 7 about here

4. DISCUSSION

4.1 Conclusions and Theoretical Implications

Designing and implementing an effective service delivery system presupposes the understanding and assessment of perceived service complexity, especially as perceived from frontline staff whose role is central in delivering the service offering to customers (Barnes *et al.* 2011). Whereas prior research advances operational, internal supplier or structural assessments of service complexity (e.g. Coelho *et al.*, 2011; Braun and Hadwich, 2016), this study advances a conceptualization of PSC from a FLE perspective. A theoretically supported conceptualization of PSC is proposed, uncovering the multi-dimensional formative nature of the construct, in line with prior work which views complexity in different settings (e.g. market complexity) as a formative construct (Diamantopoulos *et al.* 2008). The nomological validity of the PSC construct is also examined, by testing its influence on role clarity and job performance. This work adds to the service management literature and especially to the organizational frontline stream in

confirming perceived service complexity as a suppressor of FLEs' performance and in uncovering its key underlying elements.

The results of the study indicate that FLEs' perceived complexity of a service consists of three conceptually distinct elements. First, task-related complexity corresponds to FLEs' perceived complexity from the structural characteristics of a service process, including the number of tasks involved, the difficulty of tasks, the variety of tasks and the duration of tasks. The task element is emphasized from prior work in the operational management and marketing area (e.g. Aksin and Masini 2008). Second, customer participation also affects FLEs' perceived complexity of the service delivery process. Even though customers are creators of value and often viewed as 'partial employees' (e.g. Smith and Colgate 2007), in practice the integration of customers into service delivery is an arduous task (Vivek, Beatty and Morgan 2012), as extra parameters need to be added to the service system. Parameters such as the intensity of customer participation, the unpredictability of customers' behavior, the need for customization, and the simultaneous presence of many customers during service interactions can amplify FLEs' perceptions on the overall complexity of the service. This finding contributes to the customer co-destruction stream (e.g. Smith 2013; Echeverri and Skålén 2011) by identifying an additional negative consequence from intense customer participation in service encounters which impairs FLEs' ability to perform their role.

The third element of PSC captures FLEs' perceptions on the obscurity of a service. Findings advance current wisdom in setting the nature of the service as a determinant of PSC. In principle, services that are more general and mentally intangible or offer limited visibility of the overall service process are considered more complex. Both customer participation and the nature of a

service emerge as important and uncontrollable sources of perceived complexity, confirming our initial intention to depart from process-based assessments of complexity.

From a theoretical standpoint, this study contributes to the service management stream in two ways. First, it extends the job demands-resources framework in setting perceived service complexity as a job demand of FLEs’ performance. PSC emerges as a negative determinant of FLEs’ ability to perform during their interactions with customers. This is due to the higher levels of mental effort required on their behalf as well as on the additional fatigue that FLEs suffer from when high PSC is evident. This conclusion builds on recent evidence that service complexity affects other stakeholders’ (i.e. customers) cognitive effort required during a service encounter (Mikolon *et al.* 2015). Second, the job demands-resources literature is also expanded in confirming PSC as an additional job characteristic which impairs work motivation. PSC restricts service organizations’ ability to design jobs with the aim of providing greater task variety and afford considerable freedom and discretion to the FLE.

The confirmation of the construct’s criterion validity provide some additional insights for service scholars and organizational frontline research. Service complexity negatively influences FLEs’ role clarity and job performance, extending recent work in the area around the role of complexity for internal service quality (Braun and Hadwich 2016; 2017). Traditional service frameworks that view employee performance as a function of job characteristics need to account for the impact of complexity that derives from the type and nature of different service offerings. Also, some insights around customers’ disruptive impact on their exchanges with customers emerge (e.g. Chan *et al.* 2010). Customers’ participation in the service delivery process might not always be beneficial, as it increases the complexity of the process and make it harder for

FLEs to perform their role requirements. Hence, perceived service complexity is set as an important parameter of FLEs' performance during the service encounter.

4.2 Managerial Implications

Based on the prior discussion, some important managerial insights emerge for practitioners and store managers. First, an inclusive and robust measurement tool of frontline staff's PSC can be utilized to assess complexity, allowing the better management of service delivery procedures and facilitating FLEs to meet their role expectations. PSC can also prove useful for store managers to assess frontline staff's PSC and gain a more comprehensive understanding of their perceived service complexity and its sources. This construct could prove particularly useful for service organizations that deliver services of varied complexity levels (e.g. hotels, hospitals, universities) where FLEs deal with different levels of complexity given the various services they offer to customers. In such cases, managers should use our conceptualization of PSC to identify the source of perceived complexity and securely choose appropriate management practices to apply (e.g. job re-engineering, increased empowerment).

Another significant contribution of this study to service managers derives from mapping the key sources of perceived complexity for FLEs. The detailed analysis of the factors that determine PSC allows store managers to understand whether it is mostly structurally-driven, job-design related or customer-imposed. In the first case, it is recommended that service providers invest in simplifying the service delivery procedures while maintain operational efficiency. One way to do so would be to reduce the number or the difficulty of intermediate steps in the delivery process; alternatively, one could re-organize the interactions among internal resources and structures, so that the duration of the service delivery be minimized.

In the case where service delivery includes varied types of services or the nature of the service delivered enhances FLEs’ perceptions of complexity, action at the job design level should take place. Yield management can reduce the number of people being served simultaneously, or at least ensure that customers who are simultaneously present will have similar needs and preferences. On the same basis, the use of customer relationship management systems, plus customer education programmes, can ensure that customers will play their role as ‘partial employees’ adequately, ensuring that PSC, due to the intensity of their participation, will remain low. In both cases, a formal service design and the use of effective service blueprints and maps could contribute significantly to this end.

Against the mainstream practice to increasingly engage customers in service activities, firms should also examine whether customer participation in service delivery generates difficulties in FLEs’ dealing with customer demands. Incorporating FLEs’ viewpoints when designing services processes could be a first step toward this direction. FLEs’ inadequate understanding of the service is a major source of PSC. Hence, service companies are advised to invest in accommodating FLEs’ suggestions in dealing with customers and highlight the tangible parts of the service and map the intangible ones (e.g. create service maps and blueprints). At the same time, some actions need to take place for the customers’ side. For instance, task standardization (or customized standardization), together with generating adequate customer knowledge, should be adopted to reduce both the unpredictability of customers’ behavior and the customization requirements.

Following this discussion, service managers should always take into account FLEs’ perceived complexity when designing service delivery systems. More importantly, increasing service customization and empowerment strategies should not be applied arbitrarily. Also, offering a

greater service variety needs to be more carefully assessed when developing service delivery protocols, as both might impair FLEs' ability to perform their role successfully.

5. LIMITATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

Findings of the present study should be viewed in the light of certain limitations. A first important limitation pertains to the inclusion of only FLEs' views in the PSC conceptualization; as customers' roles in service delivery is also critical, their point of view could also be incorporated into the PSC scale, and future researchers could collect data from both FLEs and customers in order to understand their interplay. Second, the use of a convenience sample for these studies should render the interpretation of the findings across service settings with caution. Although several service industries and various types of FLEs were included in the samples for both studies, there is a need to gather further evidence to enable generalisability; for that reason, the PSC scale should be tested with caution in different cultural environments. This finding provides an alternative approach to future studies that might examine complexity in different circumstances and settings (e.g. manufacturing process, consumption).

Another methodological limitation pertains to the fact that the twelve sub-dimensions were captured and measured using single items. Future research should explore the possibility of these sub-dimensions being better measured by multi-item scales. This could potential further expand the conceptualisation of PSC and improve PSC scale's accuracy. Finally, another suggestion for future research pertains to the study's scope, which could be broader. The present study is mostly focused on the conceptualization and empirical validation of the PSC notion and not on an examination of the way the PSC construct interrelates with other important organizational or customer variables. Future research could develop and empirically test a conceptual framework

that incorporates PSC together with specific antecedents and consequences, both at an organizational and a customer level. The antecedents of the specific dimensions of PSC could also be examined, along with their influence on the overall PSC construct.

REFERENCES

Aksin, O. Zeynep, and Andrea Masini (2008), “Effective strategies for internal outsourcing and offshoring of business services: An empirical investigation”, *Journal of Operations Management*, 26(2), 239-256.

Amaral, Luis A. Nunes, and Brian Uzzi (2007), “Complex systems-a new paradigm for the integrative study of management, physical, and technological systems”, *Management Science*, 53(7), 1033-1035.

Andaleeb, Syed Saad, and Amiya K. Basu (1995), “Technical complexity and consumer knowledge as moderators of service quality evaluation in the automobile service industry”, *Journal of Retailing*, 70(4), 367-381.

Anderson, James C., and David W. Gerbing (1988), “Structural equation modeling in practice: A review and recommended two-step approach”, *Psychological Bulletin*, 103(3), 411-423.

Bagozzi, Richard P., and Youjae Yi (1988), “On the evaluation of structural equation models”, *Journal of the Academy of Marketing Science*, 16(1), 74-94.

Bakker, Arnold B., Evangelia Demerouti, and Martin C. Euwema (2005), “Job resources buffer the impact of job demands on burnout”, *Journal of Occupational Health Psychology*, 10(2), 170-180.

- Balaji, M. S., Sanjit Kumar Roy, and Ali Quazi (2017), "Customers' emotion regulation strategies in service failure encounters", *European Journal of Marketing*, 51(5/6), 960-982.
- Barnes, Donald C., Nicole Ponder, and Kranti Dugar (2011), "Investigating the key routes to customer delight". *Journal of Marketing Theory and Practice*, 19(4), 359-376.
- Bigné, J. Enrique, Joaquín Aldás-Manzano, Inés Küster, and Natalia Vila. (2002), "The concept mapping approach in marketing: an application in the travel agencies sector", *Qualitative Market Research: An International Journal*, 5(2), 87-95.
- Bollen, Kenneth A., and Walter R. Davis (2009), "Causal indicator models: Identification, estimation, and testing" *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 498-522.
- and Richard Lennox (1991), "Conventional wisdom on measurement: a structural equation perspective", *Psychological Bulletin*, 110(2), 305-314.
- Bradley, Graham L., Beverley A. Sparks, Dieter Zapf, Janet R. McColl-Kennedy, and Nerina L. Jimmieson (2013), "Task-Relationship-Self: A Framework for Understanding Service Encounter Behaviors", *Psychology and Marketing*, 30(6), 512-528.
- Braun, Corina, and Karsten Hadwich (2016), "Complexity of internal services: Scale development and validation", *Journal of Business Research*, 69(9), 3508-3522.
- and ----- (2017), "Determinants of perceived internal service complexity: An empirical analysis of promoting and limiting complexity factors", *European Business Review*, 29(1), 123-152.

- Brooker, Edward, Marion Joppe, Michael CG Davidson, and Kathy Marles (2012), "Innovation within the Australian outdoor hospitality parks industry", *International Journal of Contemporary Hospitality Management*, 24(5), 682-700.
- Buckley, Joan (2003), "E-service quality and the public sector", *Managing Service Quality*, 13(6), 453-462.
- Byrne, Barbara M. (2006), *Structural Equation Modeling with EQS*, 2nd Edition, Multivariate Applications Series.
- Chase, Richard B., and David A. Tansik (1983), "The customer contact model for organization design", *Management Science*, 29(9), 1037-1050.
- Chan, Kimmy Wa, Chi Kin Yim, and Simon K. Lam (2010), "Is customer participation in value creation a double-edged sword? Evidence from professional financial services across cultures", *Journal of marketing*, 74(3), 48-64.
- Chen, Gilad, Wendy J. Casper, and Jose M. Cortina (2001), "The roles of self-efficacy and task complexity in the relationships among cognitive ability, conscientiousness, and work-related performance: A meta-analytic examination", *Human Performance*, 14 (3), 209-230.
- Chen, Yang, Yi Wang, Saggi Nevo, Jiafei Jin, Luning Wang, and Wing S. Chow (2014), "IT capability and organizational performance: the roles of business process agility and environmental factors", *European Journal of Information Systems*, 23(3), 326-342.
- Chung, Beth G., and Benjamin Schneider (2002), "Serving multiple masters: Role conflict experienced by service employees" *Journal of Services Marketing* 16(1), 70-87.
- Churchill Jr, Gilbert A. (1979), "A paradigm for developing better measures of marketing constructs", *Journal of Marketing Research*, 16(1), 64-73.

- Clark, Graham, Robert Johnston, and Michael Shulver (2000), "Exploiting the service concept for service design and development", in Fitzsimmons, J.A. and Fitzsimmons, M.J. (Eds.), *New Service Design: Creating Memorable Experiences*, Sage, Thousand Oaks, CA, 71-91.
- Coltman, Tim, Timothy M. Devinney, David F. Midgley, and Sunil Venaik (2008), "Formative versus reflective measurement models: Two applications of formative measurement", *Journal of Business Research*, 61(12), 1250-1262.
- Dagger, Tracey S., Peter J. Danaher, and Brian J. Gibbs (2009), "How often versus how long: the interplay of contact frequency and relationship duration in customer-reported service relationship strength", *Journal of Service Research*, 11(4), 371-388.
- Danaher, Peter J., and Jan Mattsson (1998), "A comparison of service delivery processes of different complexity", *International Journal of Service Industry Management*, 9(1), 48-63.
- De Brentani, Ulrike (1989), "Success and failure in new industrial services", *Journal of Product Innovation Management*, 6(4), 239-258.
- de Castro Lobo, Maria Stella, Yasar A. Ozcan, Angela CM da Silva, Marcos P. Estellita Lins, and Roberto Fiszman (2010), "Financing reform and productivity change in Brazilian teaching hospitals: Malmquist approach", *Central European Journal of Operations Research*, 18(2), 141-152.
- Demerouti, Evangelia, Arnold B. Bakker, Friedhelm Nachreiner, and Wilmar B. Schaufeli (2001), "The job demands-resources model of burnout", *Journal of Applied Psychology*, 86(3), 499-512.
- Devlin, J.F. (2001), "Consumer evaluation and competitive advantage in retail financial services- A research agenda", *European Journal of Marketing*, 35(5/6), 639-660.

Devlin, James F. (2007), “Complex services and choice criteria: an example from the life assurance market”, *Journal of Marketing Management*, 23(7-8), 631-650.

Diamantopoulos, Adamantios, and Heidi M. Winklhofer (2001), “Index construction with formative indicators: an alternative to scale development”, *Journal of Marketing Research*, 38(2), 269-277.

----- and Siguaw, Judy A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, 17(4), 263-282.

-----, Petra Riefler, and Katharina P. Roth (2008), “Advancing formative measurement models”, *Journal of Business Research*, 61(12), 1203-1218.

Dong, Beibei, K. Sivakumar, Kenneth R. Evans, and Shaoming Zou (2015), "Effect of customer participation on service outcomes: The moderating role of participation readiness", *Journal of Service Research*, 18(2), 160-176.

Echeverri, Per, and Per Skålén (2011), “Co-creation and co-destruction: A practice-theory based study of interactive value formation”, *Marketing theory*, 11(3), 351-373.

Fornell, Claes, and David F. Larcker (1981), “Evaluating structural equation models with unobservable variables and measurement error”, *Journal of Marketing Research*, 18 (1), 39-50.

Germain, Richard, Cornelia Dröge, and William Christensen (2001), “The mediating role of operations knowledge in the relationship of context with performance”, *Journal of Operations Management*, 19(4), 453-469.

- Gillison, Stephanie T., William Magnus Northington, and Sharon E. Beatty (2016), "Employees' Emotional Reactions to Customer Deal Requests", *Journal of Marketing Theory and Practice*, 24(2), 147-165.
- Gounaris, Spiros, and Achilleas Boukis (2013), "The role of employee job satisfaction in strengthening customer repurchase intentions", *Journal of Services Marketing*, 27(4), 322-333.
- Hartline, Michael D., and Orville C. Ferrell (1996), "The management of customer-contact service employees: an empirical investigation", *The Journal of Marketing*, 60(4), 52-70.
- Hjort, Klas, Björn Lantz, Dag Ericsson, and John Gattorna (2013) "Customer segmentation based on buying and returning behaviour." *International Journal of Physical Distribution & Logistics Management* 43(19), 852-865.
- Hoffman, K. Douglas, and L. W. Turley (2002), "Atmospherics, service encounters and consumer decision making: an integrative perspective", *Journal of Marketing Theory and Practice*, 10(3), 33-47.
- Holm, Morten, V. Kumar, and Carsten Rohde (2012), "Measuring customer profitability in complex environments: an interdisciplinary contingency framework", *Journal of the Academy of Marketing Science*, 40(3), 387-401.
- Jan Angelis, Jannis, and Marc Thompson (2007), "Product and service complexity and high performance work practices in the aerospace industry", *Journal of Industrial Relations*, 49(5), 775-781.
- Jarvis, Cheryl Burke, Scott B. MacKenzie, and Philip M. Podsakoff (2003), "A critical review of construct indicators and measurement model misspecification in marketing and consumer research", *Journal of Consumer Research*, 30(2), 199-218.

- 1
2
3 Jinbo, H.E., Hu, B.Y., and F.A.N. Xitao, (2017), "Q-Sort Technique. In "Encyclopedia of
4
5 Personality and Individual Differences" (4), Springer International Publishing.
6
7
8 Iyer, Karthik NS, Prashant Srivastava, and Mohammed YA Rawwas (2014). "Aligning supply
9
10 chain relational strategy with the market environment: Implications for operational
11
12 performance". *Journal of Marketing Theory and practice*, 22(1), 53-72.
13
14
15 Kauppila, Olli Pekka (2014) "So, what am I supposed to do? A multilevel examination of role
16
17 clarity", *Journal of Management Studies*, 51(5), 737-763.
18
19
20 Kostopoulos, Giannis, Gounaris, Spiros and Boukis, Achilleas (2012), "Service blueprinting
21
22 effectiveness: drivers of success", *Managing Service Quality*, 22(6), 580-591.
23
24
25 Kreye, Melanie E., Jens K. Roehrich, and Michael A. Lewis (2015), "Servitising manufacturers:
26
27 the impact of service complexity and contractual and relational capabilities", *Production*
28
29 *Planning and Control*, 26(14-15), 1233-1246.
30
31
32 Lalwani, Ashok K. (2009), "The Distinct Influence of Cognitive Busyness and Need for Closure
33
34 on Cultural Differences in Socially Desirable Responding," *Journal of Consumer Research*,
35
36 36 (2), 305-316.
37
38
39 Laroche, Michel, Jasmin Bergeron, and Christine Goutaland (2001), "A three-dimensional scale
40
41 of intangibility", *Journal of Service Research*, 4(1), 26-38.
42
43
44 -----, Gordon HG McDougall, Jasmin Bergeron, and Zhiyong Yang (2004), "Exploring how
45
46 intangibility affects perceived risk", *Journal of Service Research*, 6(4), 373-389.
47
48
49 Law, Kenneth S., Chi-Sum Wong, and William M. Mobley (1998), "Toward a taxonomy of
50
51 multidimensional constructs", *Academy of Management Review*, 23(4), 741-755.
52
53
54
55
56
57
58
59
60

- Lightfoot, Howard W., and Heiko Gebauer (2011), "Exploring the alignment between service strategy and service innovation", *Journal of Service Management*, 22(5), 664-683.
- Lin, Chien-Hsin, Peter J. Sher, and Hsin-Yu Shih (2005), "Past progress and future directions in conceptualizing customer perceived value", *International Journal of Service Industry Management*, 16(4), 318-336.
- Lings, Ian N., and Roger F. Brooks (1998), "Implementing and measuring the effectiveness of internal marketing", *Journal of Marketing Management*, 14(4), 325-351.
- Lovelock, Christopher H. (1983), "Classifying services to gain strategic marketing insights", *The Journal of Marketing*, 47(3), 9-20.
- MacKenzie, Scott B., Philip M. Podsakoff, and Nathan P. Podsakoff (2011), "Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques", *MIS Quarterly*, 35(2), 293-334.
- Malhotra, Naresh K. (1981), "A scale to measure self-concepts, person concepts, and product concepts", *Journal of marketing research*, 18(4), 456-464.
- Martínez-Tur, Vicente, José M. Peiró, and José Ramos, (2001) "Linking service structural complexity to customer satisfaction: The moderating role of type of ownership." *International Journal of Service Industry Management* 12(3), 295-306.
- Mikolon, Sven, Anika Kolberg, Till Haumann, and Jan Wieseke (2015), "The Complex Role of Complexity How Service Providers Can Mitigate Negative Effects of Perceived Service Complexity When Selling Professional Services" *Journal of Service Research*, 18(4), 513-528.

Ng, Sandy, Rebekah Russell-Bennett, and Tracey Dagger (2007), “A typology of mass services: the role of service delivery and consumption purpose in classifying service experiences”, *Journal of Services Marketing*, 21(7), 471-480.

Nunnally, Jum C. (1978), *Psychometric Theory*, McGraw-Hill, New York.

Rafiq, Mohammed, and Pervaiz K. Ahmed (1998), “A customer-oriented framework for empowering service employees”, *Journal of Services Marketing*, 12(5), 379-396.

Shostack, G. Lynn (1987), “Service positioning through structural change”, *Journal of Marketing*, 51(1), 34-43.

Silvestro, Rhian, Lin Fitzgerald, Robert Johnston, and Christopher Voss (1992), “Towards a classification of service processes”, *International Journal of Service Industry Management*, 3(3), 62-75.

Simon, Françoise, and Jean-Claude Usunier (2007), “Cognitive, demographic, and situational determinants of service customer preference for personnel-in-contact over self-service technology”, *International Journal of Research in Marketing*, 24(2), 163-173.

Singal, Nidhi (2008), “Working towards inclusion: Reflections from the classroom”, *Teaching and Teacher Education*, 24(6), 1516-1529.

Singh, Jagdip (2000) “Performance productivity and quality of frontline employees in service organizations”, *Journal of Marketing*, 64(2), 15-34.

Smith, Anne (2013). “The value co-destruction process: a customer resource perspective”. *European Journal of Marketing*, 47 (11/12), 1889-1909.

Smith, J. Brock, and Mark Colgate (2007), “Customer value creation: a practical framework”. *Journal of marketing Theory and Practice*, 15(1), 7-23.

- Soteriou, Andreas C., and Richard B. Chase (1998), "Linking the customer contact model to service quality", *Journal of Operations Management*, 16(4), 495-508.
- Surprenant, Carol F., and Michael R. Solomon (1987), "Predictability and personalization in the service encounter", *Journal of Marketing*, 51 (2), 86-96.
- Swanson, Scott R., and Scott W. Kelley (2001), "Attributions and outcomes of the service recovery process". *Journal of Marketing Theory and Practice*, 9(4), 50-65.
- Teixeira, Jorge, Lia Patrício, Nuno J. Nunes, Leonel Nóbrega, Raymond P. Fisk, and Larry Constantine, (2012), "Customer experience modeling: from customer experience to service design." *Journal of Service Management* 23(3), 362-376.
- Vivek, Shiri D., Sharon E. Beatty, and Robert M. Morgan (2012), "Customer engagement: Exploring customer relationships beyond purchase", *Journal of Marketing Theory and Practice*, 20(2), 122-146.
- Vohs, Kathleen D., Roy F. Baumeister, Brandon J. Schmeichel, Jean M. Twenge, Noelle M. Nelson, and Dianne M. Tice (2008), "Making Choices Impairs Subsequent Self-Control: A Limited-Resource Account of Decision Making, Self-Regulation, and Active Initiative," *Journal of Personality and Social Psychology*, 94 (5), 883-898.
- Wang, Chung-Jen, Huei-Ting Tsai, and Ming-Tien Tsai (2014). "Linking transformational leadership and employee creativity in the hospitality industry: The influences of creative role identity, creative self-efficacy, and job complexity". *Tourism Management*, 40 (3), 79-89.
- Whitaker, Brian G., Jason J. Dahling, and Paul Levy (2007), "The development of a feedback environment and role clarity model of job performance", *Journal of Management*, 33(4), 570-591.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1: Conceptual Framework

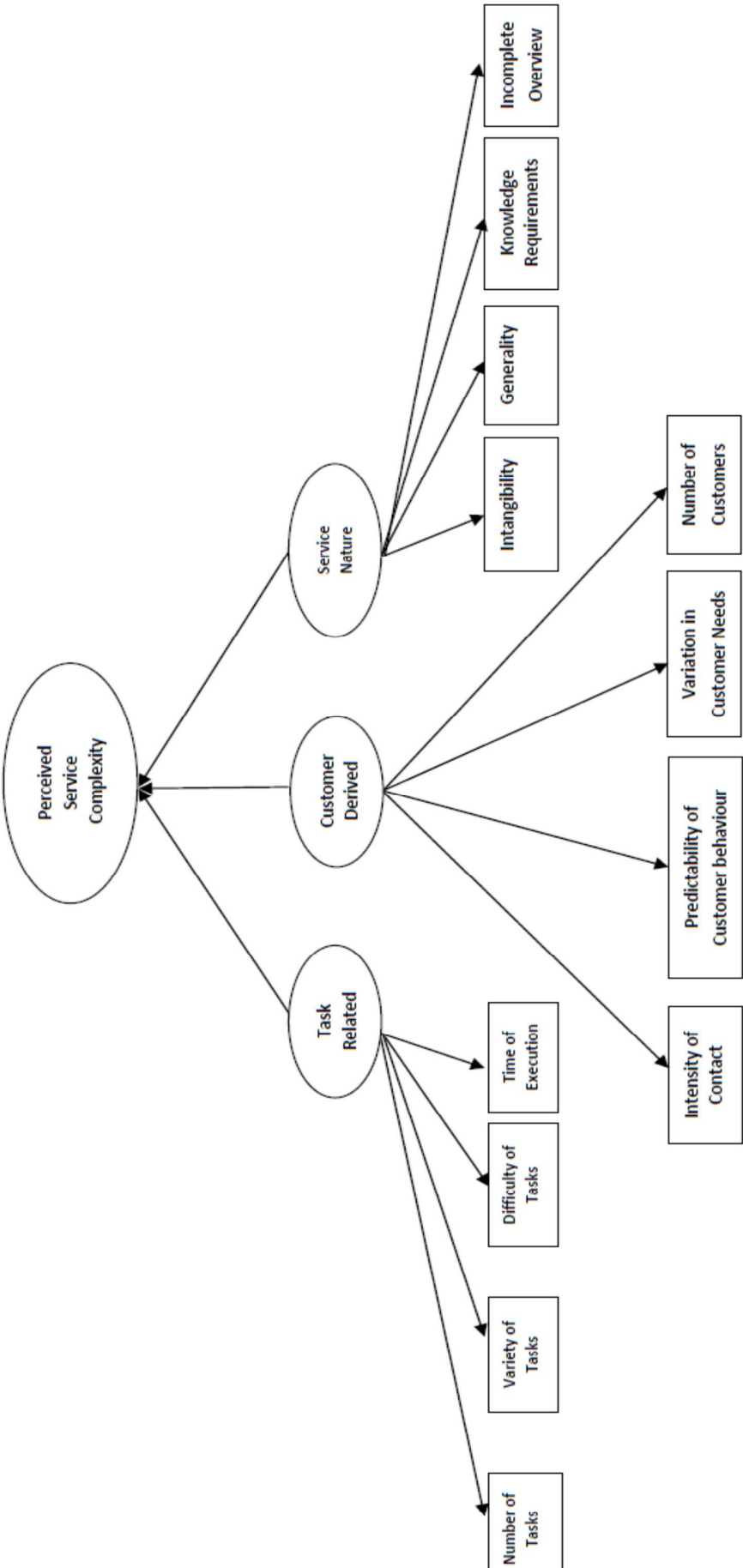


Table 1: Items' descriptive statistics and standardised factor loadings (Study 1)

Dimension	Item	Mean	Standard Deviation (SD)	CFA Loading	Item-to-Total Correlation
Task-Related Complexity	The delivery of the service involves the completion of many tasks.	4.35	1.825	0.782	0.720
	The tasks we have to execute in order to deliver the service are very different to each other.	4.26	1.716	0.798	0.736
	The tasks we have to execute in order to deliver the service are very difficult.	4.21	1.649	0.791	0.731
	It takes a lot of time to execute the tasks involved in the delivery of the service.	4.27	1.835	0.849	0.776
Customer-Derived Complexity	Customers participate intensively in the delivery of the service.	4.43	1.530	0.845	0.733
	It is difficult to predict customers' behaviour during the service delivery.	4.47	1.574	0.808	0.724
	There are many alternative ways to serve the customers, depending on their preferences.	4.67	1.561	0.608	0.552
	During our interaction with each customer, many other customers are present.	4.03	1.532	0.785	0.703
'Service Nature'-Derived Complexity	The service we provide to the customers is quite intangible.	4.51	1.370	0.602	0.527
	It requires a lot of knowledge to completely understand the service we provide to the customers.	4.58	1.448	0.746	0.651
	It is difficult for me to explain the service we provide to the customers to someone else, because it is very abstract.	4.37	1.444	0.748	0.647
	Many parts of the service delivery process are not visible to me.	4.43	1.462	0.729	0.616

Table 2: Confirmatory Factor Analysis for PSC scale (Study 1)

Factor	CFI	TLI	RMSEA	AVE	CR	Cronbach's alpha	Correlations	
							Task- Related	Customer -Derived
Task-Related	0.999	0.996	0.038	0.649	0.881	0.880		
Customer-Derived	0.999	0.997	0.030	0.588	0.849	0.842	0.358**	
'Service Nature'- Derived	0.985	0.955	0.084	0.502	0.800	0.798	0.471**	0.488**

CFI: comparative fit index; TLI: Tucker–Lewis index; RMSEA: root mean square error of approximation /
**. Correlation is significant at the 0.01 level (2-tailed). * / Correlation is significant at the 0.05 level (2-tailed).

Table 3: Items' standardised factor loadings (Study 2)

Dimension	Item	Standardised Loading	Standard Error (SE)
Task-Related Complexity	The delivery of the service involves the completion of many tasks.	0.667	0.129
	The tasks we have to execute in order to deliver the service are very different to each other.	0.510	0.136
	The tasks we have to execute in order to deliver the service are very difficult.	0.671	0.146
	It takes as a lot of time to execute the tasks involved in the delivery of the service.	0.755	-
Customer-Derived Complexity	Customers participate intensively in the delivery of the service.	0.616	0.139
	It is difficult to predict customers' behaviour during the service delivery.	0.804	0.151
	There are many alternative ways to serve the customers, depending on their preferences.	0.827	0.136
	During our interaction with each customer, many other customers are present.	0.791	-
'Service Nature'- Derived Complexity	The service we provide to the customers is quite intangible.	0.680	0.147
	It requires a lot of knowledge to completely understand the service we provide to the customers.	0.446	0.121
	It is difficult for me to explain the service we provide to the customers to someone else, because it is very abstract.	0.868	0.134
	Many parts of the service delivery process are not visible to me.	0.846	-

Table 4: Results of measurement model assessment and scale statistics (Study 2)

Factor	CR	AVE	Cronbach's alpha	Correlations	
				Task- Related	Customer -Derived
<i>Task-Related</i>					
<i>Customer-Derived</i>	0.810	0.531	0.728	0.415**	
<i>'Service Nature'-</i>	0.846	0.583	0.839	(0.17)	0.642**
<i>Derived</i>	0.817	0.527	0.742	0.557** (0.31)	(0.41)

** . Correlation is significant at the 0.01 level (2-tailed) / * Correlation is significant at the 0.05 level (2-tailed).

Table 5: Regression weights for the second-order formative model (Study 2)

Standardised Regression Weights	Estimate	SE
Task-Related Complexity → PSC	0.550	0.067**
Customer-Derived Complexity → PSC	0.423	0.056**
'Service Nature'-Derived Complexity → PSC	0.198	0.045**

** . Correlation is significant at the 0.01 level (2-tailed) / *Correlation is significant at the 0.05 level (2-tailed).

Table 6: Criterion validity test (Study 2)

<i>Measurement Model and Correlations</i>					
	CR	AVE	Cronbach's alpha	Correlations	
				Role Clarity	Job Performan ce
<i>PSC</i>	-	-	0.871	-0.640**	-0.537**
<i>Role Clarity</i>	0.904	0.654	0.904		0.694**
<i>Job Performance</i>	0.825	0.542	0.823		

** . Correlation is significant at the 0.01 level (2-tailed)/ * Correlation is significant at the 0.05 level (2-tailed).

Table 7: Standardised regression weights for criterion validity test (Study 2)

Path	Estimate	SE
PSC → Role Clarity	-0.714	0.197**
Role Clarity → Job Performance	0.426	0.120**
PSC→Job Performance	-0.543	0.244**

** . Correlation is significant at the 0.01 level (2-tailed)/ *Correlation is significant at the 0.05 level (2-tailed)

Overall comments to reviewers

we would like to thank again all three reviewers for providing such a detailed, insightful and constructive feedback on this paper. Following your recommendations, some important changes took place as follows:

- 1. The introduction section was restructured and the construct is more clearly distinct from recent work in the area which examines employees' complexity.**
- 2. The definition of PSC was added in the abstract to highlight the main contribution of the study and some clearer implications are now reported.**
- 3. Section 2.1 which explains the rational for this new conceptualisation in relation to the pertinent literature was re-written and hopefully a more robust case is now presented.**
- 4. Some inconsistent terms have been replaced and some additional info was added on the research design and sample selection.**
- 5. A number of other important changes also took place which are highlighted in our response to each reviewer.**

Reviewer(s)' Comments to Author:

Reviewer: 2

Comments to the Author

This review is for the revised version of the manuscript, whose objectives are to (1) conceptualize perceived service complexity (PSC) and its dimensions, (2) develop a valid PSC scale, and (4) "assess the criterion validity of a formative model to describe PSC" and its relationships with role clarity and job performance (pp. 3-4).

As before, the manuscript is reasonably well written. There still is a problem with the domain of the PSC scale, because the manuscript defines its domain by selecting PSC dimensions from the existing literature (section 2.2), which is not the same as developing a construct's domain with exploratory research because published studies may have missed part of the domain. Furthermore, the scale items' topics result from the lit review, rather than developing a pool of items that cover the entire PSC domain, and then choosing the best items and topics from an iterative scale purification process. This is not the best scale development procedure, and the inconsistency in the discussion emphasizes the problem. In other words, the manuscript accepting the domain as defined by the existing literature (the three dimensions) would be alright, but suggesting that the entire domain is captured easily could be inaccurate. (Defining the domain of a measured construct comes at the beginning of scale development, not at the end. See, e.g., Churchill 1979.)

Response: Thank you for this comment. We acknowledge that the current manuscript accepts the domain as defined by the existing literature. This is now clearly reflected in the paper (see p. 6, paragraphs, 2 & 3). The aim of the literature review was indeed to identify dimensions of the construct that have been documented in the literature and integrate them into a single

conceptualisation. We have amended the wording in some of the sentences in sections 2.1 and 2.2 to clarify this issue.

The explanation for why constructs are reflective or formative seems rather lengthy and defensive, but perhaps this is a response to a perceived weakness? The resultant model makes sense to me, because the three first-order constructs create a formative index of perceived complexity and do not need to be correlated. However, the first-order constructs probably also could be modeled as reflective with no loss to theory.

Response: Thank you for this comment. Including the explanation about the nature of the construct was viewed as important for three reasons. Following the discussion on the pertinent literature around developing new constructs (e.g. Treiblmaier et al. 2011; MacKenzie et al. 2011; Roy et al. 2012; Costa and Anderson 2011), most scale developments in the marketing literature are of reflective nature and thus we considered vital to justify the formative structure of this new construct. Moreover, most of the pertinent conceptualizations of service complexity include reflective constructs and therefore our differentiation should be justified. Last, it was requested from reviewers in prior submissions and it was deemed important.

On the other hand, I'm against using single items to capture dimensions, although this is hidden because the many (12) dimensions are measured by a single item each and combined to form the three first-order constructs in the PSC scale. Using single items reduces reliability (consistency), as evidenced by the existence of some low factor loadings in Study 2 (Table 3).

The reporting of the structural equation modeling results is improved, but the fit statistics continue to be marginal at best. With every model, the χ^2/df ratio is well over 3.0, and the CFI and TLI are less than the recommended 0.95, suggesting a marginal to poor model fit. In my experience, these minimum figures are not difficult to achieve, particularly as the sample sizes are not overly large (150 and 244 for studies 1 and 2, respectively). The RMSEA figures are fine (<0.08), as are the GFI figures (surprisingly), but the overall impression is that the models do not fit well because the χ^2 figure is the fundamental fit statistic on which other fit statistics are based. I don't know where is the problem: it could be the items comprising the PSC scale? It would be interesting to see the modification indices and other output for possible solutions to the question of the poor fits. I should add that much work has explored fit statistics over the years (see, e.g., Hooper, Coughlan, and Mullen (2008) in *The Electronic Journal of Business Research Methods*, or Hu and Bentler (1999) in *Structural Equation Modeling: A Multidisciplinary Journal*), so using the otherwise excellent Bollen (1989) as support for low fit figures is dated in this regard.

It is unusual and interesting that the fit statistics seemingly offer an inconsistent interpretation of fit, as the recommended χ^2/df ratio, CFI, and TLI (NNFI) figures typically are not difficult to achieve, whereas the GFI recommendation typically is very difficult to achieve. Again, estimating SEMs for the individual factors (Table 2) still doesn't make sense because there is no power to reject a false model (these models have just 2 df). Note that some of Table 3's factor loadings are quite low (0.510, 0.446), which indicates these items are not working well with Study 2's validation data.

Response: thank you for this comment as well. We acknowledge that using single items to capture the 12 dimensions is a potential limitation and that this may have influenced the moderate fit of

the final model. We have now added the following evidence to reflect that limitation (see Section 5).

“Another methodological limitation pertains to the fact that the twelve sub-dimensions were captured and measured using single items. Future research should explore the possibility of these sub-dimensions being better measured by multi-item scales.”

Regarding your comments on the fit indices, “the CFI and TLI are less than the recommended 0.95, suggesting a marginal to poor model fit”, we are aware that these indices are not the recommended ones but they are accepted from the pertinent literature (Hair et al., 1998; Ferreira and Franco, 2017). The recommended acceptance of a good fit to a model requires that the obtained normed fit index (NFI), Tucker-Lewis Index (TLI) and comparative fit index (CFI) values be greater than or equal to 0.90, while the root mean square error of approximation (RMSEA) should be below 0.08 (Medsker, et. al., 1994; Byrne, 2016; Hair, Black, Barry, and Anderson, 2010).

We were also to spot a number of studies embarking in construct development that also report similar factor loadings in one or two items of the developing construct (i.e. Tambyah et al., 2009; Yang et al., 2004; Walsh and Beatty, 2007; Karatepe et al., 2005).

Acknowledging your concerns, we would be happy to run some additional data collection in a different sample of respondents and drawing on the same context and check whether we are able to confirm the results we got from study 1. However, this would require a significant amount of time so that we can run this additional study.

References

Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge.

Costa, A. C., & Anderson, N. (2011). Measuring trust in teams: Development and validation of a multifaceted measure of formative and reflective indicators of team trust. *European Journal of Work and Organizational Psychology*, 20(1), 119-154.

Ferreira, A., & Franco, M. (2017). The Mediating Effect of Intellectual Capital in The Relationship Between Strategic Alliances and Organizational Performance in Portuguese Technology-Based SMEs. *European Management Review*, 14(3), 303-318.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate data analysis* (Vol. 5, No. 3, pp. 207-219). Upper Saddle River, NJ: Prentice hall.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Applications of SEM. *Multivariate data analysis*.

Karatepe, O. M., Yavas, U., & Babakus, E. (2005). Measuring service quality of banks: Scale development and validation. *Journal of Retailing and Consumer Services*, 12(5), 373-383.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS quarterly*, 35(2), 293-334.

Medsker, G. J., Williams, L. J., & Holahan, P. J. (1994). A review of current practices for evaluating causal models in organizational behavior and human resources management research. *Journal of management*, 20(2), 439-464.

Roy, S., Tarafdar, M., Ragu-Nathan, T. S., & Erica, M. (2012). The Effect of Misspecification of Reflective and Formative Constructs in Operations and Manufacturing Management Research. *Electronic Journal of Business Research Methods*, 10(1).

Tambyah, S. K., Tuyet Mai, N. T., & Jung, K. (2009). Measuring status orientations: Scale development and validation in the context of an Asian transitional economy. *Journal of Marketing Theory and Practice*, 17(2), 175-187.

Treiblmaier, H., Bentler, P. M., & Mair, P. (2011). Formative constructs implemented via common factors. *Structural Equation Modeling*, 18(1), 1-17.

Walsh, G., & Beatty, S. E. (2007). Customer-based corporate reputation of a service firm: scale development and validation. *Journal of the academy of marketing science*, 35(1), 127-143.

Yang, Z., Jun, M., & Peterson, R. T. (2004). Measuring customer perceived online service quality: scale development and managerial implications. *International Journal of Operations & Production Management*, 24(11), 1149-1174.

Minor Comments

p. 3 Braun and Hadwich (2017) is cited (also on p. 24) but not listed in the Reference section.

Response: our apologies for this inconsistency. The reference list and the corresponding use of them in the main body was double-checked.

pp. 14-15 Why use role clarity and job performance for the criterion (validity check) variables? Why not variables such as job satisfaction or turnover intentions? The choices could use support.

Response: thank you for stressing this out. Some additional justification was added to support the selection of these two outcome variables (see last paragraph, p.14-p.15). The following phrase was added "(..) were selected as they constitute the most immediate consequences of high complexity for FLEs while they can remain pivotal in defining customers' experience with the service encounter (Whitaker, Dahling and Levy 2007)." As the scope of including the two variables was to test the scale's criterion validity and not to advance theory on the consequences of complexity, we believe that their selection makes sense, as they have been tested previously in the literature (Kaupilla 2014).

References

Kauppila, Olli-Pekka (2014) “So, what am I supposed to do? A multilevel examination of role clarity”, *Journal of Management Studies*, 51(5), 737-763.

Whitaker, Brian G., Jason J. Dahling, and Paul Levy (2007), “The development of a feedback environment and role clarity model of job performance”, *Journal of Management*, 33(4), 570-591.

p. 17 I still don’t understand the purpose of the t-test used to show “significant differences between low and high complexity service jobs”. All this shows is that there is variation in perceptions of job complexity – how this information is useful could be explained?

Response: thank you for this comment as well. This test was added in request of another reviewer who wanted to ensure that service complexity of various service sectors included in study 1 is not pre-determined by the research team but rather, various participants also report on whether they believe that their job is complex. Although we have requested respondents to complete this questionnaire with one service in mind (if they are engage in the provision of more than one services) and also to indicate the type of this service. Running this t-test makes sense as we had to ensure whether different services whether actually perceived of varied complexity. One of the basic arguments of this paper is that FLEs’ perceptions of complexity derive from various aspects of their role (e.g. customers, nature of service) and not just from the service process itself, which might vary across services.

p. 19 I continue to believe that either (1) another conceptual model be used to illustrate the model with the criterion variables, or else (2) only the full model should be illustrated if space is a consideration. (the conceptual model has latent variables only)

Response: thank you for this comment as well. We have considered as well to illustrate the full model and concur that the full model should be presented ideally. Unfortunately there is not enough space to do so, given the current format of the manuscript (we are slightly over the suggested page limit already) and it might be arbitrary to remove some existing tables.

pp. 40-41 Why do Tables 1 and 3 contain different information about the items? (and why are some of the SEs missing in Table 3?)

Response: thank you for this comment as well. Table 1 provides info on the study 1 CFA results and factor loadings whereas Table 2 reports on study 2 CFA results and factor loadings. The standard errors missing in table 3 are due to the fact that one item from each factor is hold stable when a CFA runs in AMOS and therefore is not provided on the output table.

In study 1, we report the following indices whereas in study 2 we report on Mean, Standard Deviation, CFA loadings and item-to-Total Correlation, whereas in study 2 we report on standardized loadings and standard errors. If some additional info we would be happy to include it.

Reviewer: 1

Comments to the Author

Reviewer: 1

Thank you for the opportunity to review your research. I found the subject matter interesting and I thought the manuscript was well written and easy to follow. I also believe that taking the employee's perspective is very important in marketing research. Similar to any research there are strengths and limitations, I will focus on the latter in the hopes of helping the author(s) to publish this research. I believe the manuscript is much improved. Well done to the author(s).

However, my major concerns about the measurement of the construct remain. The main contribution of this work is a new construct...thus, defining and measuring the construct is of paramount importance. In the present form, readers are left with the opinions of 2/3 of 23 (academics and practitioners) for the building blocks of the construct. This is not to say the group is wrong in their assessment – only that the author(s) do not have the proof they need to establish validity. A simple solution to this concern is to collect data and evaluate the 36 items in totality. If the results match up with the current manuscript the author(s) will have provided convincing evidence of the arguments they have made in this paper.

Response: thank you for this comment as well. Below in table 1, we provide the items that were initially developed for each dimension (i.e. item options) as well as the item that was finally selected.

Table 1

Dimension	Item Options	Final Item selected
Task-Related Complexity		
<i>Number of tasks</i>	<ol style="list-style-type: none"> 1. A large number of steps is included in the delivery of the service. 2. A large number of tasks need to be performed so that the service is delivered to customers. 3. The delivery of the service involves the completion of many tasks. 	The delivery of the service involves the completion of many tasks.
<i>Variety of tasks</i>	<ol style="list-style-type: none"> 1. The tasks we have to execute in order to deliver the service are very different to each other. 2. The steps of the service delivery process are quite diverse. 3. The tasks we have to execute in order to deliver the service vary. 	The tasks we have to execute in order to deliver the service are very different to each other.
	1. The steps required to complete the	The tasks we have to

<i>Difficulty of tasks</i>	<p>delivery of this service are very demanding.</p> <p>2. The tasks we have to execute in order to deliver the service are very difficult.</p> <p>3. I have to put a lot of effort to complete the tasks required for the delivery of this service.</p>	<p>execute in order to deliver the service are very difficult.</p>
<i>Duration of tasks</i>	<p>1. It takes as a lot of time to execute the tasks involved in the delivery of the service.</p> <p>2. Completing each step of the service process is quite time-consuming.</p> <p>3. The duration of the tasks involved in the service delivery is lengthy.</p>	<p>It takes as a lot of time to execute the tasks involved in the delivery of the service.</p>
Customer-Derived Complexity		
<i>Intensity of customer participation</i>	<p>1. Customers participate in the delivery of the service to a great extent.</p> <p>2. Customers participate intensively in the delivery of the service.</p> <p>3. The delivery of this service cannot take place without customers' intensive participation.</p>	<p>Customers participate intensively in the delivery of the service.</p>
<i>Predictability of customer behavior</i>	<p>1. It is difficult to predict customers' behaviour during the service delivery.</p> <p>2. Customer reactions to service delivery efforts are often unpredictable.</p> <p>3. Customers often react in unexpected ways during service delivery.</p>	<p>It is difficult to predict customers' behaviour during the service delivery.</p>
<i>Variation in customer needs</i>	<p>1. Each customer has their own needs and therefore they should receive different treatment.</p> <p>2. There are many alternative ways to serve the customers, depending on their preferences.</p> <p>3. Customer preferences vary and hence there are several alternative ways of serving them.</p>	<p>There are many alternative ways to serve the customers, depending on their preferences.</p>
<i>Simultaneous presence of multiple customers during the service encounter</i>	<p>1. A large number of customers are usually present when service delivery takes place.</p> <p>2. In this role, many other customers are around when employees interact with customers.</p>	<p>During our interaction with each customer, many other customers are present.</p>

	3. During our interaction with each customer, many other customers are present.	
'Service Nature'-Derived Complexity		
<i>Mental intangibility</i>	1. The service we provide to the customers is quite intangible. 2. It is hard for one to grasp what the service this company offers is about. 3. The service we provide to our customers is quite difficult to describe.	The service we provide to the customers is quite intangible.
<i>Knowledge requirements</i>	1. One has to spend quite some time in this job, before being able to fully understand what the service is about. 2. Detailed knowledge is necessary for one to be able to fully understand the service to customers. 3. It requires a lot of knowledge to completely understand the service we provide to the customers.	It requires a lot of knowledge to completely understand the service we provide to the customers.
<i>Generality</i>	1. The service this company offers is quite abstract for customers 2. It is difficult for me to explain the service we provide to the customers to someone else, because it is very abstract. 3. The service we provide to the customer is very general.	It is difficult for me to explain the service we provide to the customers to someone else, because it is very abstract.
<i>Incomplete overview</i>	1. I don't have access to all parts of the service delivery process. 2. Many parts of the service delivery process are not visible to me. 3. Every single aspect of the service delivery process is observable by me.	Many parts of the service delivery process are not visible to me

Issues

1. The definition of PSC should be in the abstract. This is the main contribution of the work and thus should be highlighted
2. Also, in the abstract overly general statements are made such as "useful implications for academics and practitioners are presented". I recommend making specific implications and suggestions to hook the reader.

I would still prefer to see at least one specific implication in the abstract as opposed to the statement "useful implications" are provided. As a side note, I have never seen a response to

reviewer section that grouped comments together as opposed to answering them individually...it might be a better way...however, I worry that you miss responding directly to concern in some cases.

Response: thank you for this direction as well. The definition of PSC was added in the abstract to highlight the main contribution of the study and some clearer implications are now reported in the abstract.

3. Since the author(s) are introducing a “new” construct it is important that they carefully discriminate it from other related constructs. I am not sure they have done this in the current version of the paper. For example, how is this different from task complexity in the management literature? A recent services article looked at task complexity [Coelho and Augusto, Journal of Service Research 13(4) 426-438 2010]...importantly this article suggested that the factors in task complexity interact with each other which is much different than this article has it modeled...I realize that on page 4 the author(s) suggest that most studies looking at complexity “do not explore service complexity holistically” I am just not sure that the author(s) have done enough to differentiate their idea.

Response: Thank you for these comments. Following your comments, we have restructured section 2.1 (p.4 & 5). The differences from prior contributions are more clearly highlighted in the introduction now. We acknowledge the contribution of previous conceptualisations of service complexity which provide useful insights on how the construct should be conceptualised and measured following various approaches (e.g. Silvestro et al, 1992; Mikolon et all, 2011; etc.). We believe that this is clear through our reference to this work throughout the manuscript. The main aim of this study, as now outlined in the introduction is:

“Echoing these challenges, this study aims to expand prior conceptualizations of service complexity and provide an exhaustive view of perceived service complexity for the service encounter reality, by introducing an FLE-based conceptualization of complexity, namely Perceived Service Complexity (PSC)”

We also believe that this distinct conceptualisation (i.e using FLE’s perceptions on the basis of several dimensions we identified in the literature) offers an FLE-based perspective, which indicates the implications of PSQ for managing frontline staff.

4. In the section about identifying dimensions of PSC the author(s)s state “After an extensive review of the literature, three major factors underlying PSC were identified, each of which contains four dimensions.” What did this entail? Considering this is the crux of the paper more detail is required.

Response: thank you very much for this constructive comment. The literature review is now restructured and the construct is more clearly distinct from recent work in the area which examines employees’ complexity (Braun and Hadwich, 2016). The focus of this work lies on frontline staff’s complexity and especially the importance of complexity during frontline

employees' interactions with customers rather than complexity that arises from intra-organizational exchanges and its consequences on internal stakeholders' perceptions of complexity (Braun and Hadwich, 2017). Moreover, we also make clear (see section 2.1, p.6; also see section 3, p.15) that PSC was developed based on an extensive review of the current management, operations and marketing literature, as an exploratory approach was not used in the scale development process.

The extensive literature review taken draws on theories and work on service design (e.g. Shostack 1987; Soteriou and Chase 1998), customer participation (Buckley 2003), mental intangibility (Laroche et al. 2001; 2004) and the Job Demands-Resources literature (Demerouti et al., 2001). In specific, the relevant section was restructured and the following section was added:

"Following an extensive review of the concept across the organizational behaviour, services and operations literatures, three major dimensions underlying PSC emerge: Task-Related Complexity, Customer-Derived Complexity and Service Nature-Derived Complexity. These three dimensions represent the input of the three main sources of complexity for FLEs; first, the way the service is designed by the service provider (Chase and Tansik 1983); second, the input from external participants (i.e. the customers) (Dagger et al. 2009) and, third, the effect from the nature of the service itself, which cannot be adjusted by the service provider or customers (Laroche et al. 2001; 2004)." (p.6)

5. With regards to item generation – how many items were in the original pool? It seems like 16. This seems like a small number. Many of the construct papers I have read or been a part of used pools considerably bigger than the final scale.

Response: Thank you for bringing this to our attention. This part of the manuscript is indeed confusing and we do apologise for this. In the revised version of the document, this segment has been completely re-written to reflect the exact process that was followed. The process involved the creation of a pool of 36 items (3 alternative items for each of the 12 elements of the PSC scale). The pool of items was then given to the 10 executives and 13 academics, who were asked to rank the items based on the degree to which they believe the items are measuring what they intend to (content validity). With the use of Q-sort tests, the item with the best content validity is selected for each of the 12 elements and is included in the final research instrument. The process is explained in detail in the following paragraphs in the revised version of the manuscript:

"Following the review, 10 interviews with executives from several service firms (Hotels, Restaurants) and 13 with management and marketing academics is conducted, leading to the creation of an initial pool of items to capture the underlying elements of each dimension. For each of the 12 elements, three possible items is developed and each item is assigned a Likert-type scale with anchors 1-7 (Totally Disagree – Totally Agree).

The list of items, together with the study's overall subject and research objectives, were then given to the 10 executives and 13 academics, who were asked to rank the items based on the degree to which they believe the items are measuring what they intend to (content validity). With the use of Q-sort tests, the item with the best content validity is selected for each of the 12 elements and is included in the final research instrument. Some items' wording is slightly amended based on the experts' suggestions. In the end of this process, 12 items are included (see

Table 1) and 24 are excluded from the final questionnaire. Some indicative items that are excluded from the questionnaire because they had low content validity and/or are not categorized in any dimension by at least two-thirds of the experts participated (Malhotra, 1981) are: 'The tasks we have to execute in order to deliver the service vary', 'The duration of the tasks involved in the service delivery is lengthy', 'Customers' preferences vary and hence there are several alternative ways of serving them' and 'The service we provide to the customers is very general'."

The four items included in the end of the section are just some indicative items that were excluded from the final questionnaire. In table 1 (see p. 7 and p.8), we provide the items that were initially developed for each dimension (i.e. item options) as well as the item that was finally selected.

To be clear, step 1 was evaluating face validity. With a relatively small sample (23 people) of which 14 had to agree when choosing an item to represent a facet of the dimension (there is a scenario with 14 agreeing where the 10 non-academics all agree and only 4 of 13 academics agree with seems troublesome). This is where I remain most concerned. Why not collect data and let the results purify the measure? Especially, because nearly half the sample is not academics I am not sure they truly comprehend the conceptual domain of a facet. This is not to say the non-academics are not a very valuable part of your study. I think having them help you develop the 36 items which you test is valuable. But basing content validity on so few responses seems problematic. Once again the reason I am brining this up again is because it is central to your main contribution.

Response: Thank you for this comment. We acknowledge the reviewer's concerns, but as stated by the reviewer the aim of step 1 was to test the scale's (and therefore the items' too) face validity. The aim was not to purify the scale, which is something that happened in later steps of study 1. Hence, we followed the C-OAR-SE process having as aim to post hoc test the face validity of the scale and not to contribute to the definition of the construct or find new dimensions (Rossiter, 2002). Regarding the reliability of the practitioners' responses, we understand the reviewers concerns, but we are very positive that the experienced professionals that were included in the sample were capable of comprehending and elaborating the complicated constructs included in our study.

6. I am also very uncomfortable with a single item representing a conceptual domain of a construct. This is perhaps my biggest concern about the entire paper.

This remains a concern – especially for the reasons I highlight above. My concern is not only for single item measures – I have used them in my research as well and have had to justify their use. My concern is that this is a new construct...added to the fact that the single item was based as least partially on non-academic opinion I worry about the validity of the scales.

Response: We acknowledge that using single items to capture the 12 dimensions is a potential limitation and may have influenced the moderate fit of the final model. We have now added the following phrase in the limitations' section to reflect this concern:

"Another methodological limitation pertains to the fact that the 12 sub-dimensions were captured and measured by single item scales. Future research should explore the possibility of

these sub-dimensions being better measured by multi-item scales. This could potential further expand the conceptualisation of PSC and improve PSC scale's accuracy."

In support of the current approach, Bergkvist and Rossiter (2007, 2009) report empirical findings indicating that single-item (SI) measures demonstrated equally high predictive validity as MI scales. Despite the risks associated with single –item scales, there are circumstances in which single items could be legitimately employed (e.g. Diamantopoulos, Sarstedt, Fuchs, Wilczynski and Kaiser, 2012; Rossiter, 2002). Diamantopoulos et al., (2012) identify a number of conditions and suggest some conditions where single-item scales are acceptable and should be used (cf Bergkvist and Rossiter, 2009). For example, when a relatively small sample is used, when items are highly homogenous (Diamantopoulos et al., 2012), single-item measures can be legitimately used.

In this case all these conditions apply. The sample that was finally reached in study 1 remains moderate to low (N=150). Also, some items from the ones removed were highly homogeneous, conceptually speaking. For example, item 4 from task-related complexity (i.e. "It takes a lot of time to execute the tasks involved in the delivery of the service") was not conceptually distinct from an item removed where both were measuring task duration ("The duration of the tasks involved in the service delivery is lengthy").

Apart from this evidence, there is theoretical argument for using a single-item measure rather than a multiple-item measure in prior work in the area (Rossiter, 2002; Drolet and Morrison, 2001). Rossiter (2002) argues that a single-item measure is sufficient if the construct is such that in the minds of raters when two conditions exist. First, "the object of the construct is "concrete singular," meaning that it consists of one object that is easily and uniformly imagined, and second the attribute of the construct is "concrete," again meaning that it is easily and uniformly imagined.

In the PSC construct both conditions are evident. First, the majority of the elements for all three dimensions are simple, distinct from each other and unambiguous regarding the perceived meaning from the employee. For example, the number of tasks is a clearly distinct concept from the duration of tasks that the FLE needs to complete. Also, in the 'customer-derived complexity' dimension, the simultaneous presence of multiple customers during the service encounter is a clearly different concept from the predictability of customer behavior during a service encounter. The same applies to the elements of 'service nature'-derived complexity where incomplete overview of the service delivery process is fundamentally different from knowledge requirements for the service at hand.

7. What field did the academics and executives come from? Certainly, if one of the academics was from management I would highlight it as it would add credibility to your paper as it so deeply rooted in management theory.

Response: Thank you for this comment. The academics were indeed management and marketing academics. Also, some the executives worked in the hospitality industry (e.g. hotels, restaurants). This information is now added to the revised manuscript.

8. The sample for study 1 is very small. Because it is a new construct a larger sample size might be warranted

Response: Thank you for this comment. We do acknowledge that the sample size in Study 1 could be considered moderate to low. Nevertheless, a sample of 150 individuals for the first stage of a two-stage scale validation process is commonly considered adequate in the marketing literature (e.g. Hooley et al, 2005; Pervan, Bove and Johnson, 2009). Such a sample size allows for all statistical analysis to be valid, without jeopardising the power of the statistical tests and/or additional bias for sample size sensitive tests (Wolf et al, 2013). Thus, although the reviewers' comment is indeed valid, we considered that the study's results are adequately reliable, especially since the second study (n=244) confirmed its findings.

Acknowledging that this is a new construct, if the reviewer is not convinced by our justification, we are willing to carry out an additional data collection to supplement study 1's current sample. **Acknowledging** your concerns, we would be happy to run some additional data collection in a different sample of respondents and drawing on the same context and check whether we can confirm the results from study 1.

10. How exactly was the data collected?

11. Please provide much more information on where the data came from for Study 2. How was it collected? Once again, I would say the sample is small...

Response: Thank you for this comment. More information on the data collection in study 2 was indeed necessary. In the revised version the following paragraph was added, which explains the process that was followed, as well as the study's response rate. "In order to verify the properties of the PSC scale generated in the first study, as well as to test the second-order formative model, a second study is carried out using a different sample. Specifically, 619 FLEs are conducted in their workspace and asked to complete the study's questionnaire out of whom 244 agreed to do it (Response rate: 39.41%) All participants filled in the questionnaire individually at their organisation's premises. In all cases a fellow researcher was present during the completion of the questionnaire."

I remain unclear on how the data was collected in both studies. Was it a pen and paper survey? Was it done online? What does a fellow research was present mean? What type of company was it?

Response: Our apologies for not clarifying these issues earlier on. Some analytical information around the sampling process are now provided on page 16, section 3.2.1.

12. Was the single item that was asked in study 1 also asked in study 2?

Response: Thank you for this comment. The same item was addressed in both cases and it is now reported in study 2. Initially it was not reported as the mix of services included in study 2 is quite similar to the one utilized for study 1 (see p.19 for more details)

13. Did the author(s) collect any data that could account for differences in traits/skills of employees in their sample? Could these be included in statistical analyses in any manner?

Response: Actually, there is some additional data in studies 1 & 2 but it is mostly related to demographics such as educational level, training, years of working experience and tenure in the current employer. Potentially, a number of personality traits could be examined but we think that this apply to most work in the organizational frontline research area and would require a different approach and theoretical background (e.g. acknowledging FLEs' intrapersonal traits and attributes) (e.g. Di Mascio, 2010), which would totally change the focal point of this paper.

Following your recommendations, we have tested for the potential influence of demographic characteristics and work experience and we didn't identify any significant changes in the results. So, as this was not part of the scope of this research, these variables were kept out of the hypothesised models. Also, after some search in pertinent work, we have not been able to identify evidence suggesting different perceptions of complexity based on one's demographics.

14. In general, I thought the discussion was adequate.

Minor Issues

1. On page 2 the author(s) state that service deliver process is important and cite Zeithaml 1988. This is a seminal cite, but perhaps a more updated cite is warranted
2. The Aksin and Masini 2008 cite on the same page has the full name
3. Type on page 5 third sentence
4. Employees are referenced in many different ways: service employees, front-line employees, employees...maybe choose just one way
5. You probably do not need to repeat the use of AMOS in study 2.
6. On page 23 why is (JD-R) in brackets?
7. Some Journal names are not italicized
8. Missing the issue # on many references
9. I was surprised there were not more citations for JMTP as I could think of several articles in back issues that might be relevant

Response: thank for these comments as well. We have included some additional work from the journal, clarified the aforementioned issues and removed the inconsistencies mentioned.

There are grammatical errors still present in the paper. Also, the reference list has many inconsistencies on formatting – I might recommend using a source such as endnote to alleviate these issues.

Response: our apologies for these inconsistencies as well. We have thoroughly look through the references' section and any issues should be solved now.

References

- Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of marketing research*, 44(2), 175-184.
- Bergkvist, L., & Rossiter, J. R. (2009). Tailor-made single-item measures of doubly concrete constructs. *International Journal of Advertising*, 28(4), 607-621.
- Braun, Corina, and Karsten Hadwich (2016), "Complexity of internal services: Scale development and validation", *Journal of Business Research*, 69(9), 3508-3522.
- Braun, Corina, and Karsten Hadwich (2017), "Determinants of perceived internal service complexity: An empirical analysis of promoting and limiting complexity factors", *European Business Review*, 29(1), 123-152.
- Buckley, Joan (2003), "E-service quality and the public sector", *Managing Service Quality*, 13(6), 453-462.
- Demerouti, Evangelia, Arnold B. Bakker, Friedhelm Nachreiner, and Wilmar B. Schaufeli (2001), "The job demands-resources model of burnout", *Journal of Applied Psychology*, 86(3), 499-512.
- Di Mascio, R. (2010). The service models of frontline employees. *Journal of Marketing*, 74(4), 63-80.
- Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, 40(3), 434-449.
- Drolet, A. L., & Morrison, D. G. (2001). Do we really need multiple-item measures in service research?. *Journal of service research*, 3(3), 196-204.
- Hooley, G. J., Greenley, G. E., Cadogan, J. W., & Fahy, J. (2005). The performance impact of marketing resources. *Journal of business research*, 58(1), 18-27.
- Laroche, Michel, Jasmin Bergeron, and Christine Goutaland (2001), "A three-dimensional scale of intangibility", *Journal of Service Research*, 4(1), 26-38.
- Laroche, Michel, Gordon HG McDougall, Jasmin Bergeron, and Zhiyong Yang (2004), "Exploring how intangibility affects perceived risk", *Journal of Service Research*, 6(4), 373-389.
- Mikolon, Sven, Anika Kolberg, Till Haumann, and Jan Wieseke (2015), "The Complex Role of Complexity How Service Providers Can Mitigate Negative Effects of Perceived Service Complexity When Selling Professional Services" *Journal of Service Research*, 18(4), 513-528.
- Pervan, S. J., Bove, L. L., & Johnson, L. W. (2009). Reciprocity as a key stabilizing norm of interpersonal marketing relationships: Scale development and validation. *Industrial Marketing Management*, 38(1), 60-70.
- Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International journal of research in marketing*, 19(4), 305-335.
- Shostack, G.Lynn (1987), "Service positioning through structural change", *Journal of Marketing*, 51(1), 34-43.
- Silvestro, Rhian, Lin Fitzgerald, Robert Johnston, and Christopher Voss (1992), "Towards a classification of service processes", *International Journal of Service Industry Management*, 3(3), 62-75.

Soteriou, Andreas C., and Richard B. Chase (1998), "Linking the customer contact model to service quality", *Journal of Operations Management*, 16(4), 495-508.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and psychological measurement*, 73(6), 913-934.

Reviewer: 3

Comments to the Author

I like the direction of this manuscript but there are still issues to be resolved. I think the front end of the manuscript needs some work and will limit my comment to that area.

Please give the reader an inkling as to what findings and implications will be discussed in the abstract. It will likely entice them to read further. Telling the reader that "useful implications" are presented is a little too generic. I would expect every manuscript published in JMTP to have some useful implications so that phrase doesn't tell me anything about this research.

The authors allude to points made in other research without explaining why those points relate to this research. For example, in the first paragraph on page 2, the authors tell us that it is critical for key stakeholders to understand the perceived complexity of the service delivery. Who are the key stakeholders? Is it only the FLEs or is it also the service provider and the customer. If it is all, an explanation of how perceived complexity affects each would be helpful to understanding why this research is important.

Response: thank you for these comments. We fully echo your suggestions and therefore, a sentence was added in the abstract which summarizes the study's main contribution. The introduction was also adjusted to better justify the importance of service complexity for different stakeholders as well as why a frontline employee perspective is needed in the literature.

In the next paragraph, what does reduced cognitive capacity mean? What happens when an FLE has reduced cognitive capacity? Why does it reduce their commitment and to what is the commitment directed? Is it the customer? The provider?

Response: thank you for this comment as well. This argument was also restructured to explain how increased service complexity results in increase resource demands from the individual.

For your perusal, when individuals evaluate a highly complex service, they need to exert strong mental effort, as complex mental tasks require large amounts of information processing capacity (Schellekens et al. 2000). When confronted with a complex task, people do not immediately give up, but initially exert even more effort through eagerness to accomplish the task. In line with this

reasoning, this additional effort will lead to even greater cognitive demands, and the individual's cognitive capacity will decrease further (Vohs et al. 2008).

The conservation of resources theory posits that individuals try to avoid losing resources and only spend cognitive resources if the expenditure is worth the effort (Hobfoll 2002). Accordingly, upon approaching their cognitive limit, people will not completely use up their processing capacity but will switch to less effortful processing methods (Jacoby 1984). Using these simpler heuristics helps to save cognitive resources and recover (Hadjimarcou and Hu 1999). For instance, individuals in complex situations may not even try to compare alternatives or may ignore information by "tuning it out" (Jacoby, Speller, and Kohn Berning 1974), as when they focus on the size of monthly installments and ignore interest rates when deciding on a mortgage.

On page 4, the authors write that "existing definitions of service complexity provide a diverse but narrow view of the construct's meaning." What is that view and what does it mean for this research? The reader needs to know this without going and reading another article first.

At the bottom of page 4, the authors state that "existing conceptualizations of service complexity can be misleading" because they don't focus on FLEs. On whom are they focused? If FLEs have been ignored, then this is again a great opportunity to bolster the case supporting the need for this research. If these issues can be addressed, then the balance of the manuscript will make much more sense to me. It will provide a strong framework for why this construct is valuable.

Response: thank you for raising these issues. These phrases were removed and a better case is provided on the gaps that current definitions of service complexity address and on why a frontline employee-based perspective is needed (see p.4 and p.5 of the revised manuscript). The differences from other service complexity definitions are better highlighted in this section.

References

Schellekens, J.M.H., Sijtsma, G.J., Vegter, E., Meijman, T.F., 2000. Immediate and delayed after-effects of long lasting mentally demanding work. *Biologicalpsychology* 53, 37–56.

Vohs, Kathleen D., Roy F. Baumeister, Brandon J. Schmeichel, Jean M. Twenge, Noelle M. Nelson, and Dianne M. Tice (2008), "Making Choices Impairs Subsequent Self-Control: A Limited-Resource Account of Decision Making, Self-Regulation, and Active Initiative," *Journal of Personality and Social Psychology*, 94 (5), 883-898.

Hadjimarcou, J., & Hu, M. Y. (1999). Global product stereotypes and heuristic processing: The impact of ambient task complexity. *Psychology & Marketing*, 16(7), 583-612.

Jacoby, J., Speller, D. E., & Berning, C. K. (1974). Brand choice behavior as a function of information load: Replication and extension. *Journal of consumer research*, 1(1), 33-42.

Jacoby, J. (1984). Perspectives on information overload. *Journal of consumer research*, 10(4), 432-435.